(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/078516 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US02/10421

(22) International Filing Date: 28 March 2002 (28.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/280,255	30 March 2001 (30.03.2001)	US
60/315,563	28 August 2001 (28.08.2001)	US
60/347,313	9 January 2002 (09.01.2002)	US

(71) Applicant *(for all designated States except US)*: **CORIXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WANG, Tongtong** [US/US]; 8049 N.E. 28th Street, Medina, WA 98039 (US). **WANG, Siqing (aka "Steven")** [US/US]; 10145 224th Avenue N.E., Redmond, WA 98053 (US). **BANGUR, Chaitanya, S.** [IN/US]; 2102 N. 105th Street, Apt. J101, Seattle, WA 98133 (US). **GAIGER, Alexander** [AT/US]; 1421 42nd Avenue E., Seattle, WA 98112 (US).

(74) Agents: **CHRISTIANSEN, William, T.**; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 et al. (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF CANCER

(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, particularly cancers expressing either C-T or C-P antigens, are disclosed. Illustrative compositions comprise one or more tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly cancers expressing either C-T or C-P antigens.

WO 02/078516 A2

COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF CANCER

STATEMENT REGARDING SEQUENCE LISTING

The Sequence Listing associated with this application is provided on CD-ROM in lieu of a paper copy under Al § 801(a), and is hereby incorporated by reference into the specification. Four CD-ROMs are provided containing identical copies of the sequence listing: CD-ROM No. 1 is labeled "COPY 1 - SEQUENCE LISTING PART," contains the file 565pc.app.txt which is 2.0 MB and created on March 28 2001; CD-ROM No.2 is labeled "COPY 2 - SEQUENCE LISTING," contains the file 565pc.app.txt which is 2.0 MB and created on March 28, 2002; CD-ROM No. 3 is labeled "COPY 3 - SEQUENCE LISTING PART," contains the file 565pc.app.txt which is 2.0 MB and created on March 28, 2002; CD-ROM No. 4 is labeled "CRF Copy," contains the file 565pc.app.txt which is 2.0 Mb and created on March 28, 2002.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates generally to therapy and diagnosis of cancer. The invention is more specifically related to polypeptides, comprising at least a portion of a tumor protein, and to polynucleotides encoding such polypeptides. Such polypeptides and polynucleotides are useful in pharmaceutical compositions, *e.g.*, vaccines, and other compositions for the diagnosis and treatment of cancer.

Description of Related Art

Cancer is a significant health problem throughout the world. Although advances have been made in detection and therapy of cancer, no vaccine or other universally successful method for prevention and/or treatment is currently available. Current therapies, which are generally based on a

combination of chemotherapy or surgery and radiation, continue to prove inadequate in many patients.

However, recent progress has been made in the identification of tumor associated antigens, which are capable of eliciting cytotoxic T lymphocytes (CTL) responses. Many of these antigens are expressed at high levels in various tumors of distinct histological types, whereas their expression is silent in normal tissues with the exception of testis and placenta. Based on the mRNA expression pattern of these genes, they are defined as either cancer-testis (CT) genes or cancer-placenta (CP) genes. Due to the highly restricted expression of these CT and CP antigens in a variety of tumors types it is likely that these genes encode antigens that will constitute useful targets in the detection, prevention, and treatment of cancer.

In spite of considerable research into therapies for cancer it remains difficult to diagnose and treat effectively. Accordingly, there is a need in the art for improved methods for detecting and treating such cancers. The present invention fulfills these needs and further provides other related advantages.

BRIEF SUMMARY OF THE INVENTION

In one aspect, the present invention provides polynucleotide compositions comprising a sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(b) complements of the sequences provided in SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(d) sequences that hybridize to a sequence provided in SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911, under moderately stringent conditions;

(e) sequences having at least 75% identity to a sequence of SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(f) sequences having at least 90% identity to a sequence of SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911; and

(g) degenerate variants of a sequence provided in SEQ ID NO: 1-2785, 2787-2894, 2900-2901, and 2904-2911.

In another aspect, the present invention provides polypeptide compositions comprising a sequence selected from the group consisting of:

(a) sequences encoded by a polynucleotide sequence provided in SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(b) sequences having at least 70% identity to a sequence encoded by a polynucleotide sequence provided in SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(c) sequences having at least 90% identity to a sequence encoded by a polynucleotide sequence provided in SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911;

(d) sequences provided in SEQ ID NOs:2786, 2895-2899, 2902-2903, and 2912-2919;

(e) sequences having at least 70% identity to sequences provided in SEQ ID NOs: 2786, 2895-2899, 2902-2903, and 2912-2919; and

(f) sequences having at least 90% identity to a sequence provided in SEQ ID NOs:2786, 2895-2899, 2902-2903, and 2912-2919.

In one preferred embodiment, the polynucleotide compositions of the invention are expressed in at least about 20%, more preferably in at least about 30%, and most preferably in at least about 50% of tumors samples tested, at a level that is at least about 2-fold, preferably at least about 5-fold, and most preferably at least about 10-fold higher than that for normal tissues.

The present invention, in another aspect, provides polypeptide compositions comprising an amino acid sequence that is encoded by a polynucleotide sequence described above.

The present invention further provides polypeptides compositions comprising an amino acid sequence selected from the group consisting of sequences recited in SEQ ID NOs: 2786, 2895-2899, 2902-2903, and 2912-2919.

In certain preferred embodiments, the polypeptides and/or polynucleotides of the present invention are immunogenic, *i.e.*, they are capable of eliciting an immune response, particularly a humoral and/or cellular immune response, as further described herein.

The present invention further provides fragments, variants and/or derivatives of the disclosed polypeptide and/or polynucleotide sequences, wherein the fragments, variants and/or derivatives preferably have a level of immunogenic activity of at least about 50%, preferably at least about 70% and more preferably at least about 90% of the level of immunogenic activity of a polypeptide sequence set forth in SEQ ID NOs: 2786, 2895-2899, 2902-2903, and 2912-2919 or a polypeptide sequence encoded by a polynucleotide sequence set forth in SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911.

The present invention further provides polynucleotides that encode a polypeptide described above, expression vectors comprising such polynucleotides and host cells transformed or transfected with such expression vectors.

Within other aspects, the present invention provides pharmaceutical compositions comprising a polypeptide or polynucleotide as described above and a physiologically acceptable carrier.

Within a related aspect of the present invention, the pharmaceutical compositions, *e.g.*, vaccine compositions, are provided for prophylactic or therapeutic applications. Such compositions generally comprise an immunogenic polypeptide or polynucleotide of the invention and an immunostimulant, such as an adjuvant.

The present invention further provides pharmaceutical compositions that comprise: (a) an antibody or antigen-binding fragment

thereof that specifically binds to a polypeptide of the present invention, or a fragment thereof; and (b) a physiologically acceptable carrier.

Within further aspects, the present invention provides pharmaceutical compositions comprising: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) a pharmaceutically acceptable carrier or excipient. Illustrative antigen presenting cells include dendritic cells, macrophages, monocytes, fibroblasts and B cells.

Within related aspects, pharmaceutical compositions are provided that comprise: (a) an antigen presenting cell that expresses a polypeptide as described above and (b) an immunostimulant.

The present invention further provides, in other aspects, fusion proteins that comprise at least one polypeptide as described above, as well as polynucleotides encoding such fusion proteins, typically in the form of pharmaceutical compositions, *e.g.*, vaccine compositions, comprising a physiologically acceptable carrier and/or an immunostimulant. The fusions proteins may comprise multiple immunogenic polypeptides or portions/variants thereof, as described herein, and may further comprise one or more polypeptide segments for facilitating the expression, purification and/or immunogenicity of the polypeptide(s).

Within further aspects, the present invention provides methods for stimulating an immune response in a patient, preferably a T cell response in a human patient, comprising administering a pharmaceutical composition described herein. The patient may be afflicted with a cancer expressing either CT or CP mRNA, in which case the methods provide treatment for the disease, or a patient considered at risk for such a disease may be treated prophylactically.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient a pharmaceutical composition as recited above. The patient may be afflicted with a cancer expressing either CT or CP mRNA, in which case the

methods provide treatment for the disease, or patient considered at risk for such a disease may be treated prophylactically.

The present invention further provides, within other aspects, methods for removing tumor cells from a biological sample, comprising contacting a biological sample with T cells that specifically react with a polypeptide of the present invention, wherein the step of contacting is performed under conditions and for a time sufficient to permit the removal of cells expressing the protein from the sample.

Within related aspects, methods are provided for inhibiting the development of a cancer in a patient, comprising administering to a patient a biological sample treated as described above.

Methods are further provided, within other aspects, for stimulating and/or expanding T cells specific for a polypeptide of the present invention, comprising contacting T cells with one or more of: (i) a polypeptide as described above; (ii) a polynucleotide encoding such a polypeptide; and/or (iii) an antigen presenting cell that expresses such a polypeptide; under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells. Isolated T cell populations comprising T cells prepared as described above are also provided.

Within further aspects, the present invention provides methods for inhibiting the development of a cancer in a patient, comprising administering to a patient an effective amount of a T cell population as described above.

The present invention further provides methods for inhibiting the development of a cancer in a patient, comprising the steps of: (a) incubating $CD4^+$ and/or $CD8^+$ T cells isolated from a patient with one or more of: (i) a polypeptide comprising at least an immunogenic portion of polypeptide disclosed herein; (ii) a polynucleotide encoding such a polypeptide; and (iii) an antigen-presenting cell that expressed such a polypeptide; and (b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient. Proliferated cells may, but need not, be cloned prior to administration to the patient.

Within further aspects, the present invention provides methods for determining the presence or absence of a cancer in a patient comprising: (a) contacting a biological sample obtained from a patient with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; and (c) comparing the amount of polypeptide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within preferred embodiments, the binding agent is an antibody, more preferably a monoclonal antibody.

The present invention also provides, within other aspects, methods for monitoring the progression of a cancer in a patient. Such methods comprise the steps of: (a) contacting a biological sample obtained from a patient at a first point in time with a binding agent that binds to a polypeptide as recited above; (b) detecting in the sample an amount of polypeptide that binds to the binding agent; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polypeptide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

The present invention further provides, within other aspects, methods for determining the presence or absence of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample a level of a polynucleotide, preferably mRNA, that hybridizes to the oligonucleotide; and (c) comparing the level of polynucleotide that hybridizes to the oligonucleotide with a predetermined cut-off value, and therefrom determining the presence or absence of a cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a

polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the progression of a cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with an oligonucleotide that hybridizes to a polynucleotide that encodes a polypeptide of the present invention; (b) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; (c) repeating steps (a) and (b) using a biological sample obtained from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Within further aspects, the present invention provides antibodies, such as monoclonal antibodies, that bind to a polypeptide as described above, as well as diagnostic kits comprising such antibodies. Diagnostic kits comprising one or more oligonucleotide probes or primers as described above are also provided.

These and other aspects of the present invention will become apparent upon reference to the following detailed description. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

SEQ ID NOs:1-14 were isolated from a cDNA library constructed from pooled placenta and testis mRNA. These sequences are described in detail Table 2.

SEQ ID NOs:15-21 were isolated from the S1 subtraction library. These sequences are described in more detail in Table 3.

SEQ ID NOs:22-193 were isolated from the S2 subtraction library. These sequences are described in more detail in Table 4.

SEQ ID NOs:194-525 were isolated from the S2 subtraction library. These sequences showed no significant homology to any known sequences.

SEQ ID NOs:526-570 are group III cDNAs which demonstrated good electronic expression profiles.

SEQ ID NOs:571-2784 are group III cDNAs which demonstrated good expression profiles as determined by electronic Northern analysis.

SEQ ID NO:2785 discloses the full length cDNA sequence of MMP11.

SEQ ID NO:2786 discloses the full length amino acid sequence for MMP11.

SEQ ID NOs:2787-2889 were isolated using eSubtraction and eNorthern. These sequences were shown to have a tumor:normal ratio of greater than 5.

SEQ ID NO: 2890 discloses the full-length cDNA sequence of the GAGE-2 protein.

SEQ ID NO: 2891 discloses the full-length cDNA sequence of the Testis Specific Protein Y.

SEQ ID NO:2892 discloses the full length cDNA sequence of the putative cell surface receptor NYD-SP8 protein.

SEQ ID NO: 2893 discloses the full-length cDNA sequence of the Mitogen responsive phosphoprotein DOC-2 protein.

SEQ ID NO: 2894 discloses the full-length cDNA sequence of the Desmoglein protein.

SEQ ID NO: 2895 discloses the full-length amino acid sequence of the GAGE-2 protein.

SEQ ID NO: 2896 discloses the full-length amino acid sequence of the Testis Specific Protein Y.

SEQ ID NO:2897 discloses the full-length amino acid sequence of the putative cell surface receptor NYD-SP8 protein.

SEQ ID NO: 2898 discloses the full-length amino acid sequence of the Mitogen responsive phosphoprotein DOC-2 protein.

SEQ ID NO: 2899 discloses the full-length amino acid sequence of the Desmoglein protein.

SEQ ID NO:2900 discloses the cDNA sequence for a splice variant of desmocollin-1.

SEQ ID NO:2901 discloses the cDNA sequence for a splice variant of desmocollin-1.

SEQ ID NO:2902 discloses the amino acid sequence for a splice variant of desmocollin-1 described in SEQ ID NO:2900.

SEQ ID NO:2903 discloses the amino acid sequence for a splice variant of desmocollin-1 described in SEQ ID NO:2901.

SEQ ID NO:2904 is the full length cDNA sequence for the matrix metalloproteinase 10 gene.

SEQ ID NO:2905 is the full length cDNA sequence for the matrix metalloproteinase 1 gene.

SEQ ID NO:2906 is the full length cDNA sequence for the matrix metalloproteinase 13 gene.

SEQ ID NO:2907 is the full length cDNA sequence for the matrix metalloproteinase 24 gene.

SEQ ID NO:2908 is the full length cDNA sequence for the matrix metalloproteinase 7 gene.

SEQ ID NO:2909 is the full length cDNA sequence for the matrix metalloproteinase 26 gene.

SEQ ID NO:2910 is the full length cDNA sequence for the matrix metalloproteinase 12 gene.

SEQ ID NO:2911 is the full length cDNA sequence for the matrix metalloproteinase 3 gene.

SEQ ID NO:2912 is the full length amino acid sequence for the matrix metalloproteinase 1 gene.

SEQ ID NO:2913 is the full length amino acid sequence for the matrix metalloproteinase 10 gene.

SEQ ID NO:2914 is the full length amino acid sequence for the matrix metalloproteinase 13 gene.

SEQ ID NO:2915 is the full length amino acid sequence for the matrix metalloproteinase 24 gene.

SEQ ID NO:2916 is the full length amino acid sequence for the matrix metalloproteinase 7 gene.

SEQ ID NO:2917 is the full length amino acid sequence for the matrix metalloproteinase 26 gene.

SEQ ID NO:2918 is the full length amino acid sequence for the matrix metalloproteinase 12 gene.

SEQ ID NO:2919 is the full length amino acid sequence for the matrix metalloproteinase 3 gene.

SEQ ID NO:2920 is the cDNA sequence of clone KIAA1755.

SEQ ID NO:2921 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2922 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2923 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2924 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2925 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2926 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2927 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2928 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2929 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2930 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2931 is an amino acid sequence derived from clone KIAA1755.

SEQ ID NO:2932 is a cDNA sequence for clone OF1096C or 182036.1 (SEQ ID NO:577).

SEQ ID NO:2933 is a cDNA sequence for clone 94818 identified from the TPS1 subtraction library.

SEQ ID NO:2934 is a cDNA sequence for clone 94819 identified from the TPS1 subtraction library.

SEQ ID NO:2935 is a cDNA sequence for clone 94820 identified from the TPS1 subtraction library.

SEQ ID NO:2936 is a cDNA sequence for clone 94821 identified from the TPS1 subtraction library.

SEQ ID NO:2937 is a cDNA sequence for clone 94822 identified from the TPS1 subtraction library.

SEQ ID NO:2938 is a cDNA sequence for clone 94823 identified from the TPS1 subtraction library.

SEQ ID NO:2939 is a cDNA sequence for clone 94824 identified from the TPS1 subtraction library.

SEQ ID NO:2940 is a cDNA sequence for clone 94825 identified from the TPS1 subtraction library.

SEQ ID NO:2941 is a cDNA sequence for clone 94826 identified from the TPS1 subtraction library.

SEQ ID NO:2942 is a cDNA sequence for clone 94827 identified from the TPS1 subtraction library.

SEQ ID NO:2943 is a cDNA sequence for clone 94828 identified from the TPS1 subtraction library.

SEQ ID NO:2944 is a cDNA sequence for clone 94829 identified from the TPS1 subtraction library.

SEQ ID NO:2945 is a cDNA sequence for clone 94830 identified from the TPS1 subtraction library.

SEQ ID NO:2946 is a cDNA sequence for clone 94831 identified from the TPS1 subtraction library.

SEQ ID NO:2947 is a cDNA sequence for clone 94832 identified from the TPS1 subtraction library.

SEQ ID NO:2948 is a cDNA sequence for clone 94833 identified from the TPS1 subtraction library.

SEQ ID NO:2949 is a cDNA sequence for clone 94834 identified from the TPS1 subtraction library.

SEQ ID NO:2950 is a cDNA sequence for clone 94835 identified from the TPS1 subtraction library.

SEQ ID NO:2951 is a cDNA sequence for clone 94836 identified from the TPS1 subtraction library.

SEQ ID NO:2952 is a cDNA sequence for clone 94837 identified from the PPS1 subtraction library.

SEQ ID NO:2953 is a cDNA sequence for clone 94838 identified from the PPS1 subtraction library.

SEQ ID NO:2954 is a cDNA sequence for clone 94839 identified from the PPS1 subtraction library.

SEQ ID NO:2955 is a cDNA sequence for clone 94840 identified from the PPS1 subtraction library.

SEQ ID NO:2956 is a cDNA sequence for clone 94842 identified from the PPS1 subtraction library.

SEQ ID NO:2957 is a cDNA sequence for clone 94844 identified from the PPS1 subtraction library.

SEQ ID NO:2958 is a cDNA sequence for clone 94908 identified from the PPS1 subtraction library.

SEQ ID NO:2959 is a cDNA sequence for clone 94909 identified from the PPS1 subtraction library.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and their use in the therapy and diagnosis of cancer, particularly cancers expressing either cancer-testis (CT) antigens or cancer-prostate (CP) antigens. The importance of these antigens has emerged lately with the discovery of several families of genes, which are expressed only in tumor and not in any normal tissue except testis and placenta. These families include the MAGE family (De Smet et al., *Cancer Res. 58(4)*:743-52 (1998); Inoue at al., *Int. J. Cancer 63(4)*:523-26 (1995)); the GAGE family (De Backer et al., *Cancer Res. 59(13)*:3157-65 (1999); Brinkmann et al., *Cancer Res. 59(7)*:1445-48 (1999); Van den Eynde et al., *J. Exp. Med. 182(3)*:689-98 (1995)); and the BAGE family (Boel et al., *Immunity 2(2)*:167-75 (1995)). These genes have been detected in various tumors of different histological types, including melanoma (Kirkin et al., *APMIS 106(7)*:665-79 (1998)), lung (Scanlan et al., *Cancer Lett. 150(2)*:155-64 (2000)), esophageal (Inoue at al., *Int. J. Cancer 63(4)*:523-26 (1995)), head and neck tumors (Van den Eynde et al., *J. Exp. Med. 182(3)*:689-98 (1995)), and bladder tumors (Van den Eynde et al., *J. Exp. Med. 182(3)*:689-98 (1995)). As these genes are expressed on a wide variety of tumor types and are highly restricted in their expression on normal tissue, *i.e.*, testis and placenta, these antigens may constitute useful targets for immunotherapy. As described further below, illustrative compositions of the present invention include, but are not restricted to, polypeptides, particularly immunogenic polypeptides, polynucleotides encoding such polypeptides, antibodies and other binding agents, antigen presenting cells (APCs) and immune system cells (*e.g.*, T cells).

The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the

skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. Molecular Cloning: A Laboratory Manual (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, A Practical Guide to Molecular Cloning (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

Polypeptide Compositions

As used herein, the term "polypeptide" " is used in its conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. A polypeptide may be an entire protein, or a subsequence thereof. Particular polypeptides of interest in the context of this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

Particularly illustrative polypeptides of the present invention comprise those polypeptide sequences set forth in any one of SEQ ID NOs:

2786, 2895-2899, 2902-2903, and 2912-2919 or those encoded by a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911, or a sequence that hybridizes under moderately stringent conditions, or, alternatively, under highly stringent conditions, to a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911. Certain other illustrative polypeptides of the invention comprise amino acid sequences set forth in anyone of SEQ ID NOs: 2786, 2895-2899, 2902-2903, and 2912-2919 or polypeptides encoded by any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911.

The polypeptides of the present invention are sometimes herein referred to as CT proteins or CP proteins or CT polypeptides or CP polypeptides, as an indication that their identification has been based at least in part upon their increased levels of expression in tumor samples as wells as either testis or placenta. Thus, a CT or CP tumor polypeptide" or "CT or CP tumor protein," refers generally to a polypeptide sequence of the present invention, or a polynucleotide sequence encoding such a polypeptide, that is expressed in a substantial proportion of tumor samples, for example preferably greater than about 20%, more preferably greater than about 30%, and most preferably greater than about 50% or more of tumor samples tested, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in normal tissues, as determined using a representative assay provided herein. A CT or CP tumor polypeptide sequence of the invention, based upon its increased level of expression in tumor cells, has particular utility both as a diagnostic marker as well as a therapeutic target, as further described below.

In certain preferred embodiments, the polypeptides of the invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens

can be performed using methods such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies detected using, for example, ^{125}I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides disclosed herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically reactive (*i.e.,* specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology,* 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.,* they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g.,* in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, *e.g.,* having greater than about 100% or 150% or more immunogenic activity.

In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a polypeptide composition of the invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

In another embodiment of the invention, polypeptides are provided that comprise one or more polypeptides that are capable of eliciting T cells and/or antibodies that are immunologically reactive with one or more polypeptides described herein, or one or more polypeptides encoded by contiguous nucleic acid sequences contained in the polynucleotide sequences disclosed herein, or immunogenic fragments or variants thereof, or to one or more nucleic acid sequences which hybridize to one or more of these sequences under conditions of moderate to high stringency.

The present invention, in another aspect, provides polypeptide fragments comprising at least about 5, 10, 15, 20, 25, 50, or 100 contiguous amino acids, or more, including all intermediate lengths, of a polypeptide compositions set forth herein, such as polypeptides set forth in SEQ ID NOs: 2786, 2895-2899, 2902-2903, and 2912-2919 or those encoded by a polynucleotide sequence set forth in a sequence of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911.

In another aspect, the present invention provides variants of the polypeptide compositions described herein. Polypeptide variants generally encompassed by the present invention will typically exhibit at least about 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more identity (determined as described below), along its length, to a polypeptide sequences set forth herein.

In one preferred embodiment, the polypeptide fragments and variants provide by the present invention are immunologically reactive with an antibody and/or T-cell that reacts with a full-length polypeptide specifically set for the herein.

In another preferred embodiment, the polypeptide fragments and variants provided by the present invention exhibit a level of immunogenic activity of at least about 50%, preferably at least about 70%, and most preferably at least about 90% or more of that exhibited by a full-length polypeptide sequence specifically set forth herein.

A polypeptide "variant," as the term is used herein, is a polypeptide that typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their immunogenic activity as described herein and/or using any of a number of techniques well known in the art.

For example, certain illustrative variants of the polypeptides of the invention include those in which one or more portions, such as an N-terminal leader sequence or transmembrane domain, have been removed. Other illustrative variants include variants in which a small portion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids) has been removed from the N- and/or C-terminal of the mature protein.

In many instances, a variant will contain conservative substitutions. A "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the polynucleotides and polypeptides of the present invention and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is

desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, immunogenic variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

Table 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity

and charge characteristics (Kyte and Doolittle, 1982). These values are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (–0.4); threonine (–0.7); serine (–0.8); tryptophan (–0.9); tyrosine (–1.3); proline (–1.6); histidine (–3.2); glutamate (–3.5); glutamine (–3.5); aspartate (–3.5); asparagine (–3.5); lysine (–3.9); and arginine (–4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 ± 1); glutamate (+3.0 ± 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (–0.4); proline (–0.5 ± 1); alanine (–0.5); histidine (–0.5); cysteine (–1.0); methionine (–1.3); valine (–1.5); leucine (–1.8); isoleucine (–1.8); tyrosine (–2.3); phenylalanine (–2.5); tryptophan (–3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl- methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydropathic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be "identical" if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS 5*:151-153; Myers, E.W. and Muller W. (1988) *CABIOS 4*:11-17; Robinson, E.D. (1971) *Comb. Theor 11*:105; Saitou, N. Nei, M. (1987) *Mol. Biol. Evol. 4*:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San

Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA 80*:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or

deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide may be a fusion polypeptide that comprises multiple polypeptides as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence, such as a known tumor protein. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide, allowing the production of increased levels, relative to a non-fused polypeptide, in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene 40*:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA 83*:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide as described herein together with an unrelated immunogenic protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see*, for example, Stoute et al. *New Engl. J. Med.*, *336*:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a Mycobacterium sp., such as a Mycobacterium tuberculosis-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid. MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been described (for example, U.S. Patent Application 60/158,585; see also, Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). C-terminal fragments of the MTB32A coding sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous immunogenic polypeptides with which it is fused. One preferred Ra12 fusion polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about 90%

identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium Haemophilus influenza B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemaglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene 43*:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology 10*:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention, when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced in vivo stimulation of $CD4^+$ T-cells specific for the polypeptide.

Polypeptides of the invention are prepared using any of a variety of well known synthetic and/or recombinant techniques, the latter of which are further described below. Polypeptides, portions and other variants generally less than about 150 amino acids can be generated by synthetic means, using techniques well known to those of ordinary skill in the art. In one illustrative example, such polypeptides are synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain. *See* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146, 1963. Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division (Foster City, CA), and may be operated according to the manufacturer's instructions.

In general, polypeptide compositions (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

Polynucleotide Compositions

The present invention, in other aspects, provides polynucleotide compositions. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free

of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

As will be understood by those skilled in the art, the polynucleotide compositions of this invention can include genomic sequences, extra-genomic and plasmid-encoded sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides, peptides and the like. Such segments may be naturally isolated, or modified synthetically by the hand of man.

As will be also recognized by the skilled artisan, polynucleotides of the invention may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules may include HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a polypeptide/protein of the invention or a portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence.

Therefore, according to another aspect of the present invention, polynucleotide compositions are provided that comprise some or all of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911, complements of a polynucleotide sequence

set forth in any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911, and degenerate variants of a polynucleotide sequence set forth in any one of SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911. In certain preferred embodiments, the polynucleotide sequences set forth herein encode immunogenic polypeptides, as described above.

In other related embodiments, the present invention provides polynucleotide variants having substantial identity to the sequences disclosed herein in SEQ ID NOs: 1-2785, 2787-2894, 2900-2901, and 2904-2911, for example those comprising at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher, sequence identity compared to a polynucleotide sequence of this invention using the methods described herein, (*e.g.*, BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

Typically, polynucleotide variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompasses homologous genes of xenogenic origin.

In additional embodiments, the present invention provides polynucleotide fragments comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 10, 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length between the quoted values, such as 16, 17, 18, 19, *etc.*; 21, 22, 23, *etc.*; 30, 31,

32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through 200-500; 500-1,000, and the like.

In another embodiment of the invention, polynucleotide compositions are provided that are capable of hybridizing under moderate to high stringency conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50ºC-60ºC, 5 X SSC, overnight; followed by washing twice at 65ºC for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60-65°C or 65-70°C.

In certain preferred embodiments, the polynucleotides described above, *e.g.*, polynucleotide variants, fragments and hybridizing sequences, encode polypeptides that are immunologically cross-reactive with a polypeptide sequence specifically set forth herein. In other preferred embodiments, such polynucleotides encode polypeptides that have a level of immunogenic activity of at least about 50%, preferably at least about 70%, and more preferably at least about 90% of that for a polypeptide sequence specifically set forth herein.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may

be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

When comparing polynucleotide sequences, two sequences are said to be "identical" if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS 5*:151-153; Myers, E.W. and Muller W. (1988) *CABIOS 4*:11-17; Robinson, E.D. (1971) *Comb. Theor 11*:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol. 4*:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman

Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA 80*:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid bases occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

It will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

Therefore, in another embodiment of the invention, a mutagenesis approach, such as site-specific mutagenesis, is employed for the preparation of immunogenic variants and/or derivatives of the polypeptides described herein. By this approach, specific modifications in a polypeptide sequence can be made through mutagenesis of the underlying polynucleotides that encode them. These techniques provides a straightforward approach to prepare and test

sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the polynucleotide.

Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Mutations may be employed in a selected polynucleotide sequence to improve, alter, decrease, modify, or otherwise change the properties of the polynucleotide itself, and/or alter the properties, activity, composition, stability, or primary sequence of the encoded polypeptide.

In certain embodiments of the present invention, the inventors contemplate the mutagenesis of the disclosed polynucleotide sequences to alter one or more properties of the encoded polypeptide, such as the immunogenicity of a polypeptide vaccine. The techniques of site-specific mutagenesis are well-known in the art, and are widely used to create variants of both polypeptides and polynucleotides. For example, site-specific mutagenesis is often used to alter a specific portion of a DNA molecule. In such embodiments, a primer comprising typically about 14 to about 25 nucleotides or so in length is employed, with about 5 to about 10 residues on both sides of the junction of the sequence being altered.

As will be appreciated by those of skill in the art, site-specific mutagenesis techniques have often employed a phage vector that exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are readily commercially-available and their use is generally well-known to those skilled in the art. Double-stranded plasmids are also routinely employed in site directed mutagenesis that eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double-stranded vector that includes within its sequence a DNA sequence that encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis provides a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants. Specific details regarding these methods and protocols are found in the teachings of Maloy *et al.*, 1994; Segal, 1976; Prokop and Bajpai, 1991; Kuby, 1994; and Maniatis *et al.*, 1982, each incorporated herein by reference, for that purpose.

As used herein, the term "oligonucleotide directed mutagenesis procedure" refers to template-dependent processes and vector-mediated propagation which result in an increase in the concentration of a specific nucleic acid molecule relative to its initial concentration, or in an increase in the concentration of a detectable signal, such as amplification. As used herein, the term "oligonucleotide directed mutagenesis procedure" is intended to refer to a process that involves the template-dependent extension of a primer molecule. The term template dependent process refers to nucleic acid synthesis of an

RNA or a DNA molecule wherein the sequence of the newly synthesized strand of nucleic acid is dictated by the well-known rules of complementary base pairing (see, for example, Watson, 1987). Typically, vector mediated methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by U.S. Patent No. 4,237,224, specifically incorporated herein by reference in its entirety.

In another approach for the production of polypeptide variants of the present invention, recursive sequence recombination, as described in U.S. Patent No. 5,837,458, may be employed. In this approach, iterative cycles of recombination and screening or selection are performed to "evolve" individual polynucleotide variants of the invention having, for example, enhanced immunogenic activity.

In other embodiments of the present invention, the polynucleotide sequences provided herein can be advantageously used as probes or primers for nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence disclosed herein will find particular utility. Longer contiguous identical or complementary sequences, *e.g.*, those of about 20, 30, 40, 50, 100, 200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are also envisioned, such as the use of the sequence information for the preparation of mutant species primers, or primers for use in preparing other genetic constructions.

Polynucleotide molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide sequence disclosed herein, are particularly contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. This would allow a gene product, or fragment thereof, to be analyzed, both in diverse cell types and also in various bacterial cells. The total size of fragment, as well as the size of the complementary stretch(es), will ultimately depend on the intended use or application of the particular nucleic acid segment. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 15 and about 100 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

The use of a hybridization probe of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid molecules having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Hybridization probes may be selected from any portion of any of the sequences disclosed herein. All that is required is to review the sequences set forth herein, or to any continuous portion of the sequences, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a probe or primer. The choice of probe and primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence.

Small polynucleotide segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical

means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such as the PCR™ technology of U.S. Patent 4,683,202 (incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

The nucleotide sequences of the invention may be used for their ability to selectively form duplex molecules with complementary stretches of the entire gene or gene fragments of interest. Depending on the application envisioned, one will typically desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template, less stringent (reduced stringency) hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ salt conditions such as those of from about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus,

hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

According to another embodiment of the present invention, polynucleotide compositions comprising antisense oligonucleotides are provided. Antisense oligonucleotides have been demonstrated to be effective and targeted inhibitors of protein synthesis, and, consequently, provide a therapeutic approach by which a disease can be treated by inhibiting the synthesis of proteins that contribute to the disease. The efficacy of antisense oligonucleotides for inhibiting protein synthesis is well established. For example, the synthesis of polygalactauronase and the muscarine type 2 acetylcholine receptor are inhibited by antisense oligonucleotides directed to their respective mRNA sequences (U.S. Patent 5,739,119 and U.S. Patent 5,759,829). Further, examples of antisense inhibition have been demonstrated with the nuclear protein cyclin, the multiple drug resistance gene (MDG1), ICAM-1, E-selectin, STK-1, striatal $GABA_A$ receptor and human EGF (Jaskulski et al., *Science 240*(4858):1544-46 (1988); Vasanthakumar and Ahmed, *Cancer Commun. 1*(4):225-32 (1989); Peris et al., *Brain Res. Mol. Brain Res. 57*(2):310-20 (1998); U.S. Patent Nos. 5,801,154; 5,789,573; 5,718,709 and 5,610,288). Antisense constructs have also been described that inhibit and can be used to treat a variety of abnormal cellular proliferations, *e.g.*, cancer (U.S. Patent Nos. 5,747,470; 5,591,317 and 5,783,683).

Therefore, in certain embodiments, the present invention provides oligonucleotide sequences that comprise all, or a portion of, any sequence that is capable of specifically binding to polynucleotide sequence described herein, or a complement thereof. In one embodiment, the antisense oligonucleotides comprise DNA or derivatives thereof. In another embodiment, the oligonucleotides comprise RNA or derivatives thereof. In a third embodiment, the oligonucleotides are modified DNAs comprising a phosphorothioated modified backbone. In a fourth embodiment, the oligonucleotide sequences comprise peptide nucleic acids or derivatives thereof. In each case, preferred compositions comprise a sequence region that is complementary, and more

preferably substantially-complementary, and even more preferably, completely complementary to one or more portions of polynucleotides disclosed herein. Selection of antisense compositions specific for a given gene sequence is based upon analysis of the chosen target sequence and determination of secondary structure, T_m, binding energy, and relative stability. Antisense compositions may be selected based upon their relative inability to form dimers, hairpins, or other secondary structures that would reduce or prohibit specific binding to the target mRNA in a host cell. Highly preferred target regions of the mRNA, are those which are at or near the AUG translation initiation codon, and those sequences which are substantially complementary to 5' regions of the mRNA. These secondary structure analyses and target site selection considerations can be performed, for example, using v.4 of the OLIGO primer analysis software and/or the BLASTN 2.0.5 algorithm software (Altschul et al., *Nucleic Acids Res. 25*(17):3389-402 (1997)).

The use of an antisense delivery method employing a short peptide vector, termed MPG (27 residues), is also contemplated. The MPG peptide contains a hydrophobic domain derived from the fusion sequence of HIV gp41 and a hydrophilic domain from the nuclear localization sequence of SV40 T-antigen (Morris et al., *Nucleic Acids Res. 25*(14):2730-36 (1997)). It has been demonstrated that several molecules of the MPG peptide coat the antisense oligonucleotides and can be delivered into cultured mammalian cells in less than 1 hour with relatively high efficiency (90%). Further, the interaction with MPG strongly increases both the stability of the oligonucleotide to nuclease and the ability to cross the plasma membrane.

According to another embodiment of the invention, the polynucleotide compositions described herein are used in the design and preparation of ribozyme molecules for inhibiting expression of the tumor polypeptides and proteins of the present invention in tumor cells. Ribozymes are RNA-protein complexes that cleave nucleic acids in a site-specific fashion. Ribozymes have specific catalytic domains that possess endonuclease activity (Kim and Cech, *Proc. Natl. Acad. Sci. USA 84*(24):8788-92 (1987); Forster and

Symons, *Cell 49*(2):211-20 (1987)). For example, a large number of ribozymes accelerate phosphoester transfer reactions with a high degree of specificity, often cleaving only one of several phosphoesters in an oligonucleotide substrate (Cech et al., *Cell 27*(3 Pt 2):487-96 (1981); Michel and Westhof, *J. Mol. Biol. 216*(3):585-610 (1990); Reinhold-Hurek and Shub, *Nature 357*(6374):173-76 (1992)). This specificity has been attributed to the requirement that the substrate bind via specific base-pairing interactions to the internal guide sequence ("IGS") of the ribozyme prior to chemical reaction.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site

of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woolf et al., *Proc. Natl. Acad. Sci. USA 89*(16):7305-09 (1992)). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi et al., *Nucleic Acids Res. 20*(17):4559-65 (1992). Examples of hairpin motifs are described by Hampel et al. (Eur. Pat. Appl. Publ. No. EP 0360257), Hampel and Tritz, *Biochemistry 28*(12):4929-33 (1989); Hampel et al., *Nucleic Acids Res. 18*(2):299-304 (1990), and U.S. Patent No. 5,631,359. An example of the hepatitis δ virus motif is described by Perrotta and Been, *Biochemistry 31*(47):11843-52 (1992); an example of the RNaseP motif is described by Guerrier-Takada et al., *Cell 35*(3 Pt 2):849-57 (1983); Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, *Cell 61*(4):685-96 (1990); Saville and Collins, *Proc. Natl. Acad. Sci. USA 88*(19):8826-30 (1991); Collins and Olive, *Biochemistry 32*(11):2795-99 (1993)); and an example of the Group I intron is described in (U.S. Patent No. 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

Ribozymes may be designed as described in Int. Pat. Appl. Publ. No. WO 93/23569 and Int. Pat. Appl. Publ. No. WO 94/02595, each specifically incorporated herein by reference) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. While

specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see *e.g.*, Int. Pat. Appl. Publ. Nos. WO 92/07065, WO 93/15187, and WO 91/03162; Eur. Pat. Appl. Publ. No. 92110298.4; U. S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan et al. (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular, subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Int. Pat. Appl. Publ. Nos. WO 94/02595 and WO 93/23569, each specifically incorporated herein by reference.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II

(pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells Ribozymes expressed from such promoters have been shown to function in mammalian cells. Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

In another embodiment of the invention, peptide nucleic acids (PNAs) compositions are provided. PNA is a DNA mimic in which the nucleobases are attached to a pseudopeptide backbone (Good and Nielsen, *Antisense Nucleic Acid Drug Dev.* 7(4):431-37 (1997)). PNA is able to be utilized in a number methods that traditionally have used RNA or DNA. Often PNA sequences perform better in techniques than the corresponding RNA or DNA sequences and have utilities that are not inherent to RNA or DNA. A review of PNA including methods of making, characteristics of, and methods of using, is provided by Corey (*Trends Biotechnol. 15(6)*:224-29 (1997)). As such, in certain embodiments, one may prepare PNA sequences that are complementary to one or more portions of the ACE mRNA sequence, and such PNA compositions may be used to regulate, alter, decrease, or reduce the translation of ACE-specific mRNA, and thereby alter the level of ACE activity in a host cell to which such PNA compositions have been administered.

PNAs have 2-aminoethyl-glycine linkages replacing the normal phosphodiester backbone of DNA (Nielsen et al., *Science 254*(5037):1497-500 (1991); Hanvey *et al.*, Science *258*(5087):1481-85 (1992); Hyrup and Nielsen, *Bioorg. Med. Chem. 4*(1):5-23 (1996)). This chemistry has three important consequences: firstly, in contrast to DNA or phosphorothioate oligonucleotides, PNAs are neutral molecules; secondly, PNAs are achiral, which avoids the

need to develop a stereoselective synthesis; and thirdly, PNA synthesis uses standard Boc or Fmoc protocols for solid-phase peptide synthesis, although other methods, including a modified Merrifield method, have been used.

PNA monomers or ready-made oligomers are commercially available from PerSeptive Biosystems (Framingham, MA). PNA syntheses by either Boc or Fmoc protocols are straightforward using manual or automated protocols (Norton et al., *Bioorg. Med. Chem.* 3(4):437-45 (1995)). The manual protocol lends itself to the production of chemically modified PNAs or the simultaneous synthesis of families of closely related PNAs.

As with peptide synthesis, the success of a particular PNA synthesis will depend on the properties of the chosen sequence. For example, while in theory PNAs can incorporate any combination of nucleotide bases, the presence of adjacent purines can lead to deletions of one or more residues in the product. In expectation of this difficulty, it is suggested that, in producing PNAs with adjacent purines, one should repeat the coupling of residues likely to be added inefficiently. This should be followed by the purification of PNAs by reverse-phase high-pressure liquid chromatography, providing yields and purity of product similar to those observed during the synthesis of peptides.

Modifications of PNAs for a given application may be accomplished by coupling amino acids during solid-phase synthesis or by attaching compounds that contain a carboxylic acid group to the exposed N-terminal amine. Alternatively, PNAs can be modified after synthesis by coupling to an introduced lysine or cysteine. The ease with which PNAs can be modified facilitates optimization for better solubility or for specific functional requirements. Once synthesized, the identity of PNAs and their derivatives can be confirmed by mass spectrometry. Several studies have made and utilized modifications of PNAs (for example, Norton et al., *Bioorg. Med. Chem.* 3(4):437-45 (1995); Petersen et al., *J. Pept. Sci.* 1(3):175-83 (1995); Orum et al., *Biotechniques* 19(3):472-80 (1995); Footer et al., *Biochemistry* 35(33):10673-79 (1996); Griffith et al., *Nucleic Acids Res.* 23(15):3003-08 (1995); Pardridge et al., *Proc. Natl. Acad. Sci. USA* 92(12):5592-96 (1995);

Boffa et al., *Proc. Natl. Acad. Sci. USA* 92(6):1901-05 (1995); Gambacorti-Passerini et al., *Blood* 88(4):1411-17 (1996); Armitage et al., *Proc. Natl. Acad. Sci. USA* 94(23):12320-25 (1997); Seeger et al., *Biotechniques* 23(3):512-7 (1997)). U.S. Patent No. 5,700,922 discusses PNA-DNA-PNA chimeric molecules and their uses in diagnostics, modulating protein in organisms, and treatment of conditions susceptible to therapeutics.

Methods of characterizing the antisense binding properties of PNAs are discussed in Rose (*Anal. Chem.* 65(24):3545-49 (1993)) and Jensen et al. (*Biochemistry* 36(16):5072-77 (1997)). Rose uses capillary gel electrophoresis to determine binding of PNAs to their complementary oligonucleotide, measuring the relative binding kinetics and stoichiometry. Similar types of measurements were made by Jensen et al. using BIAcore™ technology.

Other applications of PNAs that have been described and will be apparent to the skilled artisan include use in DNA strand invasion, antisense inhibition, mutational analysis, enhancers of transcription, nucleic acid purification, isolation of transcriptionally active genes, blocking of transcription factor binding, genome cleavage, biosensors, *in situ* hybridization, and the like.

Polynucleotide Identification, Characterization and Expression

Polynucleotides compositions of the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified, as described in more detail below, by screening a microarray of cDNAs for tumor-associated expression (*i.e.*, expression that is at least two fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al.,

Proc. Natl. Acad. Sci. USA 93:10614-10619 (1996); and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155 (1997)). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR™, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCR ™ amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl.

Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.,* a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences.

For hybridization techniques, a partial sequence may be labeled (*e.g.,* by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (see Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones.

The resulting overlapping sequences can then assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (see Triglia et al., *Nucl. Acids Res. 16*:8186 (1988)), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic. 1*:111-19 (1991)) and walking PCR (Parker et al., *Nucl. Acids. Res. 19*:3055-60 (1991)). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In other embodiments of the invention, polynucleotide sequences or fragments thereof which encode polypeptides of the invention, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of a polypeptide in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of polypeptide activity, it may be useful to encode a chimeric protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the

polypeptide-encoding sequence and the heterologous protein sequence, so that the polypeptide may be cleaved and purified away from the heterologous moiety.

Sequences encoding a desired polypeptide may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al., *Nucl. Acids Res. Symp. Ser.* 215-223 (1980), Horn, T. et al., *Nucl. Acids Res. Symp. Ser.* 225-232 (1980)). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of a polypeptide, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al., *Science 269*:202-204 (1995)) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer, Palo Alto, CA).

A newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (*e.g.*, Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a desired polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described,

for example, in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York. N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions—which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the PBLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPORT1 plasmid (Gibco BRL, Gaithersburg, MD) and the like may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are generally preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding a polypeptide, vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

In bacterial systems, any of a number of expression vectors may be selected depending upon the use intended for the expressed polypeptide. For example, when large quantities are needed, for example for the induction of

antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the polypeptide of interest may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of .beta.-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster, *J. Biol. Chem. 264*:5503-09 (1989)); and the like. pGEX Vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al., *Methods Enzymol. 153*:516-44 (1987).

In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N., *EMBO J. 6*:307-11 (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al., *EMBO J. 3*:1671-80 (1984); Broglie, R. et al., *Science 224*:838-43 (1984); and Winter, J. et al., *Results Probl. Cell Differ. 17*:85-105 (1991)). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs,

S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express a polypeptide of interest. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al., *Proc. Natl. Acad. Sci. 91*:3224-27 (1994)).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T., *Proc. Natl. Acad. Sci. 81*:3655-59 (1984)). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be

provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al., *Results Probl. Cell Differ. 20*:125-62 (1994)).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation. glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, COS, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell lines which stably express a polynucleotide of interest may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex

virus thymidine kinase (Wigler, M. et al., *Cell 11*:223-32 (1977)) and adenine phosphoribosyltransferase (Lowy, I. et al., *Cell 22*:817-23 (1990)) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al., *Proc. Natl. Acad. Sci. 77*:3567-70 (1980)); npt, which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al., *J. Mol. Biol. 150*:1-14 (1981)); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan, *Proc. Natl. Acad. Sci. 85*:8047-51 (1988)). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al., *Methods Mol. Biol. 55*:121-31 (1995)).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay

techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (*J. Exp. Med. 158*:1211-16 (1983)).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the

vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen. San Diego, Calif.) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (*Prot. Exp. Purif. 3*:263-81 (1992)) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (*DNA Cell Biol. 12*:441-53 (1993)).

In addition to recombinant production methods, polypeptides of the invention, and fragments thereof, may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J., *J. Am. Chem. Soc. 85*:2149-54 (1963)). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Alternatively, various fragments may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Antibody Compositions, Fragments Thereof and Other Binding Agents

According to another aspect, the present invention further provides binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a tumor polypeptide disclosed herein, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunogically bind," and/or is "immunologically reactive" to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant K_d. See, generally, Davies et al., *Annual. Rev. Biochem.* 59:439-73 (1990).

An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are

referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs."

Binding agents may be further capable of differentiating between patients with and without a cancer, using the representative assays provided herein. For example, antibodies or other binding agents that bind to a tumor protein will preferably generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.,* blood, sera, sputum, urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of a variety of techniques known

to those of ordinary skill in the art. See, *e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976), and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient

time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "$F(ab')_2$ " fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent $V_H::V_L$ heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al., *Proc. Nat. Acad. Sci. USA 69*:2659-62 (1972); Hochman et al., *Biochem. 15*:2706-10 (1976); and Ehrlich et al., *Biochem. 19*:4091-96 (1980).

A single chain Fv ("sFv") polypeptide is a covalently linked $V_H::V_L$ heterodimer which is expressed from a gene fusion including V_H- and V_L-

encoding genes linked by a peptide-encoding linker. Huston et al., *Proc. Nat. Acad. Sci. USA 85*(16):5879-83 (1988). A number of methods have been described to discern chemical structures for converting the naturally aggregated—but chemically separated—light and heavy polypeptide chains from an antibody V region into an sFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, U.S. Patent Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Patent No. 4,946,778, to Ladner et al.

Each of the above-described molecules includes a heavy chain and a light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRS and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region. Some FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRS. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a binding-site, the CDRs are

displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al., *Nature 349*:293-99 (1991); Lobuglio et al., *Proc. Nat. Acad. Sci. USA 86*:4220-24 (1989); Shaw et al., *J. Immunol. 138*:4534-38 (1987); and Brown et al., *Cancer Res. 47*:3577-83 (1987)), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al., *Nature 332*:323-27 (1988); Verhoeyen et al., *Science 239*:1534-36 (1988); and Jones et al., *Nature 321*:522-25 (1986)), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al., *Ann. Rev. Biochem. 59*:439-73 (1990). Thus, antigen binding specificity can be preserved in a humanized

antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (*e.g.*, solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (*e.g.*, electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs

which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y, ^{123}I, ^{125}I, ^{131}I, ^{186}Re, ^{188}Re, ^{211}At, and ^{212}Bi. Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diptheria toxin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as albumins (*e.g.*, U.S. Patent No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran (*e.g.*, U.S. Patent No. 4,699,784, to Shih et al.). A carrier may also bear an agent by noncovalent

bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Patent No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, or metal oxide, radionuclide. For example, U.S. Patent No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

T Cell Compositions

The present invention, in another aspect, provides T cells specific for a tumor polypeptide disclosed herein, or for a variant or derivative thereof. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood, or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the Isolex™ System, available from Nexell Therapeutics, Inc. (Irvine, CA; see also U.S. Patent No. 5,240,856; U.S. Patent No. 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a polypeptide, polynucleotide encoding a polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide of interest. Preferably, a tumor polypeptide or polynucleotide of the invention is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a polypeptide of the present invention if the T cells specifically proliferate, secrete cytokines or kill target cells coated with the polypeptide or expressing a gene encoding the

polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen et al., *Cancer Res. 54*:1065-70 (1994). Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.,* by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a tumor polypeptide (100 ng/ml - 100 μg/ml, preferably 200 ng/ml - 25 μg/ml) for 3 - 7 days will typically result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of cytokine release (*e.g.,* TNF or IFN-γ) is indicative of T cell activation (*see* Coligan et al., Current Protocols in Immunology, vol. 1, Wiley Interscience (Greene 1998)). T cells that have been activated in response to a tumor polypeptide, polynucleotide or polypeptide-expressing APC may be $CD4^+$ and/or $CD8^+$. Tumor polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, a related donor or an unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a tumor polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a tumor polypeptide, or a short peptide corresponding to an immunogenic portion of such a polypeptide, with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a tumor polypeptide. Alternatively, one or more T cells that proliferate in the presence

of the tumor polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution.

Pharmaceutical Compositions

In additional embodiments, the present invention concerns formulation of one or more of the polynucleotide, polypeptide, T-cell and/or antibody compositions disclosed herein in pharmaceutically-acceptable carriers for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy.

It will be understood that, if desired, a composition as disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Therefore, in another aspect of the present invention, pharmaceutical compositions are provided comprising one or more of the polynucleotide, polypeptide, antibody, and/or T-cell compositions described herein in combination with a physiologically acceptable carrier. In certain preferred embodiments, the pharmaceutical compositions of the invention comprise immunogenic polynucleotide and/or polypeptide compositions of the invention for use in prophylactic and theraputic vaccine applications. Vaccine preparation is generally described in, for example, M.F. Powell and M.J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995). Generally, such compositions will comprise one or more

polynucleotide and/or polypeptide compositions of the present invention in combination with one or more immunostimulants.

It will be apparent that any of the pharmaceutical compositions described herein can contain pharmaceutically acceptable salts of the polynucleotides and polypeptides of the invention. Such salts can be prepared, for example, from pharmaceutically acceptable non-toxic bases, including organic bases (*e.g.,* salts of primary, secondary and tertiary amines and basic amino acids) and inorganic bases (*e.g.,* sodium, potassium, lithium, ammonium, calcium and magnesium salts).

In another embodiment, illustrative immunogenic compositions, *e.g.*, vaccine compositions, of the present invention comprise DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated *in situ*. As noted above, the polynucleotide may be administered within any of a variety of delivery systems known to those of ordinary skill in the art. Indeed, numerous gene delivery techniques are well known in the art, such as those described by Rolland, *Crit. Rev. Therap. Drug Carrier Systems 15*:143-198 (1998), and references cited therein. Appropriate polynucleotide expression systems will, of course, contain the necessary regulatory DNA regulatory sequences for expression in a patient (such as a suitable promoter and terminating signal). Alternatively, bacterial delivery systems may involve the administration of a bacterium (such as *Bacillus-Calmette-Guerrin*) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope.

Therefore, in certain embodiments, polynucleotides encoding immunogenic polypeptides described herein are introduced into suitable mammalian host cells for expression using any of a number of known viral-based systems. In one illustrative embodiment, retroviruses provide a convenient and effective platform for gene delivery systems. A selected nucleotide sequence encoding a polypeptide of the present invention can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to a

subject. A number of illustrative retroviral systems have been described (*e.g.*, U.S. Pat. No. 5,219,740; Miller and Rosman, *BioTechniques 7*:980-90 (1989); Miller, A. D., *Human Gene Therapy 1*:5-14 (1990); Scarpa et al., *Virology 180*:849-52 (1991); Burns et al., *Proc. Natl. Acad. Sci. USA 90*:8033-37 (1993); and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop. 3*:102-09 (1993).

In addition, a number of illustrative adenovirus-based systems have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol. 57*:267-74 (1986); Bett et al., *J. Virol. 67*:5911-21 (1993); Mittereder et al., *Human Gene Therapy 5*:717-29 (1994); Seth et al., *J. Virol. 68*:933-40 (1994); Barr et al., *Gene Therapy 1*:51-58 (1994); Berkner, K. L., *BioTechniques 6*:616-29 (1988); and Rich et al., *Human Gene Therapy 4*:461-76 (1993)).

Various adeno-associated virus (AAV) vector systems have also been developed for polynucleotide delivery. AAV vectors can be readily constructed using techniques well known in the art. See, *e.g.*, U.S. Pat. Nos. 5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 and WO 93/03769; Lebkowski et al., *Molec. Cell. Biol. 8*:3988-96 (1988); Vincent et al. (1990) Vaccines 90 (Cold Spring Harbor Laboratory Press); Carter, B. J., *Current Opinion in Biotechnology 3*:533-39 (1992); Muzyczka, N., *Current Topics in Microbiol. and Immunol. 158*:97-129 (1992); Kotin, R. M., *Human Gene Therapy 5*:793-801 (1994); Shelling and Smith, *Gene Therapy 1*:165-69 (1994); and Zhou et al., *J. Exp. Med. 179*:1867-75 (1994).

Additional viral vectors useful for delivering the polynucleotides encoding polypeptides of the present invention by gene transfer include those derived from the pox family of viruses, such as vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the novel molecules can be constructed as follows. The DNA encoding a polypeptide is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells

which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the polypeptide of interest into the viral genome. The resulting TK.sup.(-) recombinant can be selected by culturing the cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

A vaccinia-based infection/transfection system can be conveniently used to provide for inducible, transient expression or coexpression of one or more polypeptides described herein in host cells of an organism. In this particular system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide or polynucleotides of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus recombinant transcribes the transfected DNA into RNA which is then translated into polypeptide by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See, *e.g.*, Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA 87*:6743-47 (1990); Fuerst et al., *Proc. Natl. Acad. Sci. USA 83*:8122-26(1986).

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the coding sequences of interest. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an Avipox vector is particularly desirable in human and other mammalian species since members of the Avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant Avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, *e.g.*, WO 91/12882; WO 89/03429; and WO 92/03545.

Any of a number of alphavirus vectors can also be used for delivery of polynucleotide compositions of the present invention, such as those vectors described in U.S. Patent Nos. 5,843,723; 6,015,686; 6,008,035 and 6,015,694. Certain vectors based on Venezuelan Equine Encephalitis (VEE) can also be used, illustrative examples of which can be found in U.S. Patent Nos. 5,505,947 and 5,643,576.

Moreover, molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem. 268*:6866-69 (1993), and Wagner et al., *Proc. Natl. Acad. Sci. USA 89*:6099-6103 (1992), can also be used for gene delivery under the invention.

Additional illustrative information on these and other known viral-based delivery systems can be found, for example, in Fisher-Hoch et al., *Proc. Natl. Acad. Sci. USA 86*:317-21 (1989); Flexner et al., *Ann. N.Y. Acad. Sci. 569*:86-103 (1989); Flexner et al., *Vaccine 8*:17-21 (1990); U.S. Patent Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Patent No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, *Biotechniques 6*:616-27 (1988); Rosenfeld et al., *Science 252*:431-34 (1991); Kolls et al., *Proc. Natl. Acad. Sci. USA 91*:215-19 (1994); Kass-Eisler et al., *Proc. Natl. Acad. Sci. USA 90*:11498-502 (1993); Guzman et al., *Circulation 88*:2838-48 (1993); and Guzman et al., *Cir. Res. 73*:1202-07 (1993).

In certain embodiments, a polynucleotide may be integrated into the genome of a target cell. This integration may be in the specific location and orientation *via* homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the polynucleotide may be stably maintained in the cell as a separate, episomal segment of DNA. Such polynucleotide segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle. The manner in which the expression construct is delivered to a cell and where in the cell the polynucleotide remains is dependent on the type of expression construct employed.

In another embodiment of the invention, a polynucleotide is administered/delivered as "naked" DNA, for example as described in Ulmer et al., *Science 259*:1745-49 (1993), and reviewed by Cohen, *Science 259*:1691-92 (1993). The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

In still another embodiment, a composition of the present invention can be delivered via a particle bombardment approach, many of which have been described. In one illustrative example, gas-driven particle acceleration can be achieved with devices such as those manufactured by Powderject Pharmaceuticals PLC (Oxford, UK) and Powderject Vaccines Inc. (Madison, WI), some examples of which are described in U.S. Patent Nos. 5,846,796; 6,010,478; 5,865,796; 5,584,807; and EP Patent No. 0500 799. This approach offers a needle-free delivery approach wherein a dry powder formulation of microscopic particles, such as polynucleotide or polypeptide particles, are accelerated to high speed within a helium gas jet generated by a hand held device, propelling the particles into a target tissue of interest.

In a related embodiment, other devices and methods that may be useful for gas-driven needle-less injection of compositions of the present invention include those provided by Bioject, Inc. (Portland, OR), some examples of which are described in U.S. Patent Nos. 4,790,824; 5,064,413; 5,312,335; 5,383,851; 5,399,163; 5,520,639 and 5,993,412.

According to another embodiment, the pharmaceutical compositions described herein will comprise one or more immunostimulants in addition to the immunogenic polynucleotide, polypeptide, antibody, T-cell and/or APC compositions of this invention. An immunostimulant refers to essentially any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous antigen. One preferred type of immunostimulant comprises an adjuvant. Many adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, *Bortadella pertussis* or *Mycobacterium tuberculosis* derived proteins. Certain

adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, MI); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, NJ); AS-2 (SmithKline Beecham, Philadelphia, PA); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF, interleukin-2, -7, -12, and other like growth factors, may also be used as adjuvants.

Within certain embodiments of the invention, the adjuvant composition is preferably one that induces an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (*e.g.*, IFN-γ, TNFα, IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (*e.g.*, IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of a vaccine as provided herein, a patient will support an immune response that includes Th1- and Th2-type responses. Within a preferred embodiment, in which a response is predominantly Th1-type, the level of Th1-type cytokines will increase to a greater extent than the level of Th2-type cytokines. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffman, *Ann. Rev. Immunol.* 7:145-73 (1989).

Certain preferred adjuvants for eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A, together with an aluminum salt. MPL® adjuvants are available from Corixa Corporation (Seattle, WA; *see*, for example, US Patent Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Th1 response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Patent Nos. 6,008,200 and 5,856,462.

Immunostimulatory DNA sequences are also described, for example, by Sato et al., *Science 273*:352 (1996). Another preferred adjuvant comprises a saponin, such as Quil A, or derivatives thereof, including QS21 and QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); Escin; Digitonin; or *Gypsophila* or *Chenopodium quinoa* saponins . Other preferred formulations include more than one saponin in the adjuvant combinations of the present invention, for example combinations of at least two of the following group comprising QS21, QS7, Quil A, β-escin, or digitonin.

Alternatively the saponin formulations may be combined with vaccine vehicles composed of chitosan or other polycationic polymers, polylactide and polylactide-co-glycolide particles, poly-N-acetyl glucosamine-based polymer matrix, particles composed of polysaccharides or chemically modified polysaccharides, liposomes and lipid-based particles, particles composed of glycerol monoesters, etc. The saponins may also be formulated in the presence of cholesterol to form particulate structures such as liposomes or ISCOMs. Furthermore, the saponins may be formulated together with a polyoxyethylene ether or ester, in either a non-particulate solution or suspension, or in a particulate structure such as a paucilamelar liposome or ISCOM. The saponins may also be formulated with excipients such as Carbopol^{R} to increase viscosity, or may be formulated in a dry powder form with a powder excipient such as lactose.

In one preferred embodiment, the adjuvant system includes the combination of a monophosphoryl lipid A and a saponin derivative, such as the combination of QS21 and 3D-MPL® adjuvant, as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. Another particularly preferred adjuvant formulation employing QS21, 3D-MPL® adjuvant and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Another enhanced adjuvant system involves the combination of a CpG-containing oligonucleotide and a saponin derivative particularly the

combination of CpG and QS21 is disclosed in WO 00/09159. Preferably the formulation additionally comprises an oil in water emulsion and tocopherol.

Additional illustrative adjuvants for use in the pharmaceutical compositions of the invention include Montanide ISA 720 (Seppic, France), SAF (Chiron, California, United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (*e.g.,* SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Enhanzyn®) (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. Patent Application Serial Nos. 08/853,826 and 09/074,720, the disclosures of which are incorporated herein by reference in their entireties, and polyoxyethylene ether adjuvants such as those described in WO 99/52549A1.

Other preferred adjuvants include adjuvant molecules of the general formula

(I): $HO(CH_2CH_2O)_n$-A-R,

wherein, n is 1-50, A is a bond or –C(O)-, R is C_{1-50} alkyl or Phenyl C_{1-50} alkyl.

One embodiment of the present invention consists of a vaccine formulation comprising a polyoxyethylene ether of general formula (I), wherein *n* is between 1 and 50, preferably 4-24, most preferably 9; the *R* component is C_{1-50}, preferably C_4-C_{20} alkyl and most preferably C_{12} alkyl, and *A* is a bond. The concentration of the polyoxyethylene ethers should be in the range 0.1-20%, preferably from 0.1-10%, and most preferably in the range 0.1-1%. Preferred polyoxyethylene ethers are selected from the following group: polyoxyethylene-9-lauryl ether, polyoxyethylene-9-steoryl ether, polyoxyethylene-8-steoryl ether, polyoxyethylene-4-lauryl ether, polyoxyethylene-35-lauryl ether, and polyoxyethylene-23-lauryl ether. Polyoxyethylene ethers such as polyoxyethylene lauryl ether are described in the Merck index (12^{th} edition: entry 7717). These adjuvant molecules are described in WO 99/52549.

The polyoxyethylene ether according to the general formula (I) above may, if desired, be combined with another adjuvant. For example, a

preferred adjuvant combination is preferably with CpG as described in the pending UK patent application GB 9820956.2.

According to another embodiment of this invention, an immunogenic composition described herein is delivered to a host via antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may, but need not, be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects *per se* and/or to be immunologically compatible with the receiver (*i.e.,* matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Certain preferred embodiments of the present invention use dendritic cells or progenitors thereof as antigen-presenting cells. Dendritic cells are highly potent APCs (Banchereau and Steinman, *Nature 392*:245-51 (1998)) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (*see* Timmerman and Levy, *Ann. Rev. Med. 50*:507-29 (1999)). In general, dendritic cells may be identified based on their typical shape (stellate *in situ*, with marked cytoplasmic processes (dendrites) visible *in vitro*), their ability to take up, process and present antigens with high efficiency and their ability to activate naïve T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells *in vivo* or *ex vivo*, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles antigen-loaded dendritic cells (called exosomes) may be used within a vaccine (*see* Zitvogel et al., *Nature Med. 4*:594-600 (1998)).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or

fluid. For example, dendritic cells may be differentiated *ex vivo* by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNFα to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNFα, CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are conveniently categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. However, this nomenclature should not be construed to exclude all possible intermediate stages of differentiation. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcγ receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (*e.g.,* CD54 and CD11) and costimulatory molecules (*e.g.,* CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with a polynucleotide of the invention (or portion or other variant thereof) such that the encoded polypeptide, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place *ex vivo*, and a pharmaceutical composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to a patient, resulting in transfection that occurs *in vivo*. *In vivo* and *ex vivo* transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., *Immunology and Cell Biology* 75:456-60 (1997). Antigen loading of dendritic cells may be achieved by

incubating dendritic cells or progenitor cells with the tumor polypeptide, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (*e.g.*, vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell help (*e.g.*, a carrier molecule). Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

While any suitable carrier known to those of ordinary skill in the art may be employed in the pharmaceutical compositions of this invention, the type of carrier will typically vary depending on the mode of administration. Compositions of the present invention may be formulated for any appropriate manner of administration, including for example, topical, oral, nasal, mucosal, intravenous, intracranial, intraperitoneal, subcutaneous and intramuscular administration.

Carriers for use within such pharmaceutical compositions are biocompatible, and may also be biodegradable. In certain embodiments, the formulation preferably provides a relatively constant level of active component release. In other embodiments, however, a more rapid rate of release immediately upon administration may be desired. The formulation of such compositions is well within the level of ordinary skill in the art using known techniques. Illustrative carriers useful in this regard include microparticles of poly(lactide-co-glycolide), polyacrylate, latex, starch, cellulose, dextran and the like. Other illustrative delayed-release carriers include supramolecular biovectors, which comprise a non-liquid hydrophilic core (*e.g.*, a cross-linked polysaccharide or oligosaccharide) and, optionally, an external layer comprising an amphiphilic compound, such as a phospholipid (*see e.g.*, U.S. Patent No. 5,151,254 and PCT applications WO 94/20078, WO/94/23701 and WO 96/06638). The amount of active compound contained within a sustained release formulation depends upon the site of implantation, the rate and expected duration of release and the nature of the condition to be treated or prevented.

In another illustrative embodiment, biodegradable microspheres (*e.g.,* polylactate polyglycolate) are employed as carriers for the compositions of this invention. Suitable biodegradable microspheres are disclosed, for example, in U.S. Patent Nos. 4,897,268; 5,075,109; 5,928,647; 5,811,128; 5,820,883; 5,853,763; 5,814,344, 5,407,609 and 5,942,252. Modified hepatitis B core protein carrier systems. such as described in WO/99 40934, and references cited therein, will also be useful for many applications. Another illustrative carrier/delivery system employs a carrier comprising particulate-protein complexes, such as those described in U.S. Patent No. 5,928,647, which are capable of inducing a class I-restricted cytotoxic T lymphocyte responses in a host.

The pharmaceutical compositions of the invention will often further comprise one or more buffers (*e.g.,* neutral buffered saline or phosphate buffered saline), carbohydrates (*e.g.,* glucose, mannose, sucrose or dextrans), mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, bacteriostats, chelating agents such as EDTA or glutathione, adjuvants (*e.g.,* aluminum hydroxide), solutes that render the formulation isotonic, hypotonic or weakly hypertonic with the blood of a recipient, suspending agents, thickening agents and/or preservatives. Alternatively, compositions of the present invention may be formulated as a lyophilizate.

The pharmaceutical compositions described herein may be presented in unit-dose or multi-dose containers, such as sealed ampoules or vials. Such containers are typically sealed in such a way to preserve the sterility and stability of the formulation until use. In general, formulations may be stored as suspensions, solutions or emulsions in oily or aqueous vehicles. Alternatively, a pharmaceutical composition may be stored in a freeze-dried condition requiring only the addition of a sterile liquid carrier immediately prior to use.

The development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.,* oral, parenteral, intravenous, intranasal, and

intramuscular administration and formulation, is well known in the art, some of which are briefly discussed below for general purposes of illustration.

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (see, for example, Mathiowitz et al., *Nature 386*(6623):410-14 (1997); Hwang et al., *Crit. Rev. Ther. Drug Carrier Syst.* 15(3):243-84 (1998); U. S. Patent Nos. 5,641,515; 5,580,579 and 5,792,451). Tablets, troches, pills, capsules and the like may also contain any of a variety of additional components, for example, a binder, such as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compounds may be incorporated into sustained-release preparation and formulations.

Typically, these formulations will contain at least about 0.1% of the active compound or more, although the percentage of the active ingredient(s) may, of course, be varied and may conveniently be between about

1 or 2% and about 60% or 70% or more of the weight or volume of the total formulation. Naturally, the amount of active compound(s) in each therapeutically useful composition may be prepared is such a way that a suitable dosage will be obtained in any given unit dose of the compound. Factors such as solubility, bioavailability, biological half-life, route of administration, product shelf life, as well as other pharmacological considerations will be contemplated by one skilled in the art of preparing such pharmaceutical formulations, and as such, a variety of dosages and treatment regimens may be desirable.

For oral administration the compositions of the present invention may alternatively be incorporated with one or more excipients in the form of a mouthwash, dentifrice, buccal tablet, oral spray, or sublingual orally-administered formulation. Alternatively, the active ingredient may be incorporated into an oral solution such as one containing sodium borate, glycerin and potassium bicarbonate, or dispersed in a dentifrice, or added in a therapeutically-effective amount to a composition that may include water, binders, abrasives, flavoring agents, foaming agents, and humectants. Alternatively the compositions may be fashioned into a tablet or solution form that may be placed under the tongue or otherwise dissolved in the mouth.

In certain circumstances it will be desirable to deliver the pharmaceutical compositions disclosed herein parenterally, intravenously, intramuscularly, or even intraperitoneally. Such approaches are well known to the skilled artisan, some of which are further described, for example, in U.S. Patent Nos. 5,543,158; 5,641,515 and 5,399,363. In certain embodiments, solutions of the active compounds as free base or pharmacologically acceptable salts may be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions may also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations generally will contain a preservative to prevent the growth of microorganisms.

Illustrative pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions (for example, see U.S. Patent 5,466,468). In all cases the form must be sterile and must be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and/or by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

In one embodiment, for parenteral administration in an aqueous solution, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous and intraperitoneal administration. In this connection, a sterile aqueous medium that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage may be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion, (see for example, "Remington's Pharmaceutical Sciences" 15th Edition, pages 1035-1038 and 1570-1580). Some variation in dosage will necessarily occur depending on the condition of the subject being treated. Moreover, for human administration, preparations

will of course preferably meet sterility, pyrogenicity, and the general safety and purity standards as required by FDA Office of Biologics standards.

In another embodiment of the invention, the compositions disclosed herein may be formulated in a neutral or salt form. Illustrative pharmaceutically-acceptable salts include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective.

The carriers can further comprise any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions is contemplated. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

In certain embodiments, the pharmaceutical compositions may be delivered by intranasal sprays, inhalation, and/or other aerosol delivery vehicles. Methods for delivering genes, nucleic acids, and peptide compositions directly to the lungs *via* nasal aerosol sprays has been described, *e.g.,* in U.S. Patent Nos. 5,756,353 and 5,804,212. Likewise, the delivery of drugs using intranasal microparticle resins (Takenaga et al., *J. Controlled*

Release 52(1-2):81-87 (1998)) and lysophosphatidyl-glycerol compounds (U.S. Patent No. 5,725,871) are also well-known in the pharmaceutical arts. Likewise, illustrative transmucosal drug delivery in the form of a polytetrafluoroetheylene support matrix is described in U.S. Patent No. 5,780,045.

In certain embodiments, liposomes, nanocapsules, microparticles, lipid particles, vesicles, and the like, are used for the introduction of the compositions of the present invention into suitable host cells/organisms. In particular, the compositions of the present invention may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like. Alternatively, compositions of the present invention can be bound, either covalently or non-covalently, to the surface of such carrier vehicles.

The formation and use of liposome and liposome-like preparations as potential drug carriers is generally known to those of skill in the art (see for example, Lasic, *Trends Biotechnol 16*(7):307-21 (1998); Takakura, *Nippon Rinsho 56*(3):691-95 (1998); Chandran et al., *Indian J. Exp. Biol. 35*(8):801-09 (1997); Margalit, *Crit. Rev. Ther. Drug Carrier Syst. 12*(2-3):233-61 (1995); U.S. Patent Nos. 5,567,434; 5,552,157; 5,565,213; 5,738,868 and 5,795,587, each specifically incorporated herein by reference in its entirety).

Liposomes have been used successfully with a number of cell types that are normally difficult to transfect by other procedures, including T cell suspensions, primary hepatocyte cultures and PC 12 cells (Renneisen et al., *J. Biol. Chem. 265*(27):16337-42 (1990); Muller et al., *DNA Cell Biol. 9*(3):221-29 (1990)). In addition, liposomes are free of the DNA length constraints that are typical of viral-based delivery systems. Liposomes have been used effectively to introduce genes, various drugs, radiotherapeutic agents, enzymes, viruses, transcription factors, allosteric effectors and the like, into a variety of cultured cell lines and animals. Furthermore, he use of liposomes does not appear to be associated with autoimmune responses or unacceptable toxicity after systemic delivery.

In certain embodiments, liposomes are formed from phospholipids that are dispersed in an aqueous medium and spontaneously form multilamellar concentric bilayer vesicles (also termed multilamellar vesicles (MLVs).

Alternatively, in other embodiments, the invention provides for pharmaceutically-acceptable nanocapsule formulations of the compositions of the present invention. Nanocapsules can generally entrap compounds in a stable and reproducible way (see, for example, Quintanar-Guerrero et al., *Drug Dev. Ind. Pharm. 24*(12):1113-28 (1998)). To avoid side effects due to intracellular polymeric overloading, such ultrafine particles (sized around 0.1 μm) may be designed using polymers able to be degraded *in vivo*. Such particles can be made as described, for example, by Couvreur et al., *Crit. Rev. Ther. Drug Carrier Syst. 5*(1):1-20 (1988); zur Muhlen et al., *Eur. J. Pharm. Biopharm. 45*(2):149-55 (1998); Zambaux et al., *J. Controlled Release 50*(1-3):31-40 (1998); and U.S. Patent No. 5,145,684.

Cancer Therapeutic Methods

In further aspects of the present invention, the pharmaceutical compositions described herein may be used for the treatment of cancer, particularly for the immunotherapy of cancers that express either CT or CP genes. Within such methods, the pharmaceutical compositions described herein are administered to a patient, typically a warm-blooded animal, preferably a human. A patient may or may not be afflicted with cancer. Accordingly, the above pharmaceutical compositions may be used to prevent the development of a cancer or to treat a patient afflicted with a cancer. Pharmaceutical compositions and vaccines may be administered either prior to or following surgical removal of primary tumors and/or treatment such as administration of radiotherapy or conventional chemotherapeutic drugs. As discussed above, administration of the pharmaceutical compositions may be by any suitable method, including administration by intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intradermal, anal, vaginal, topical and oral routes.

Within certain embodiments, immunotherapy may be active immunotherapy, in which treatment relies on the *in vivo* stimulation of the endogenous host immune system to react against tumors with the administration of immune response-modifying agents (such as polypeptides and polynucleotides as provided herein).

Within other embodiments, immunotherapy may be passive immunotherapy, in which treatment involves the delivery of agents with established tumor-immune reactivity (such as effector cells or antibodies) that can directly or indirectly mediate antitumor effects and does not necessarily depend on an intact host immune system. Examples of effector cells include T cells as discussed above, T lymphocytes (such as $CD8^+$ cytotoxic T lymphocytes and $CD4^+$ T-helper tumor-infiltrating lymphocytes), killer cells (such as Natural Killer cells and lymphokine-activated killer cells), B cells and antigen-presenting cells (such as dendritic cells and macrophages) expressing a polypeptide provided herein. T cell receptors and antibody receptors specific for the polypeptides recited herein may be cloned, expressed and transferred into other vectors or effector cells for adoptive immunotherapy. The polypeptides provided herein may also be used to generate antibodies or anti-idiotypic antibodies (as described above and in U.S. Patent No. 4,918,164) for passive immunotherapy.

Effector cells may generally be obtained in sufficient quantities for adoptive immunotherapy by growth *in vitro*, as described herein. Culture conditions for expanding single antigen-specific effector cells to several billion in number with retention of antigen recognition *in vivo* are well known in the art. Such *in vitro* culture conditions typically use intermittent stimulation with antigen, often in the presence of cytokines (such as IL-2) and non-dividing feeder cells. As noted above, immunoreactive polypeptides as provided herein may be used to rapidly expand antigen-specific T cell cultures in order to generate a sufficient number of cells for immunotherapy. In particular, antigen-presenting cells, such as dendritic, macrophage, monocyte, fibroblast and/or B cells, may be pulsed with immunoreactive polypeptides or transfected with one

or more polynucleotides using standard techniques well known in the art. For example, antigen-presenting cells can be transfected with a polynucleotide having a promoter appropriate for increasing expression in a recombinant virus or other expression system. Cultured effector cells for use in therapy must be able to grow and distribute widely, and to survive long term *in vivo*. Studies have shown that cultured effector cells can be induced to grow in vivo and to survive long term in substantial numbers by repeated stimulation with antigen supplemented with IL-2 (see, for example, Cheever et al., *Immunological Reviews 157*:177 (1997)).

Alternatively, a vector expressing a polypeptide recited herein may be introduced into antigen presenting cells taken from a patient and clonally propagated *ex vivo* for transplant back into the same patient. Transfected cells may be reintroduced into the patient using any means known in the art, preferably in sterile form by intravenous, intracavitary, intraperitoneal or intratumor administration.

Routes and frequency of administration of the therapeutic compositions described herein, as well as dosage, will vary from individual to individual, and may be readily established using standard techniques. In general, the pharmaceutical compositions and vaccines may be administered by injection (*e.g.*, intracutaneous, intramuscular, intravenous or subcutaneous), intranasally (*e.g.*, by aspiration) or orally. Preferably, between 1 and 10 doses may be administered over a 52 week period. Preferably, 6 doses are administered, at intervals of 1 month, and booster vaccinations may be given periodically thereafter. Alternate protocols may be appropriate for individual patients. A suitable dose is an amount of a compound that, when administered as described above, is capable of promoting an anti-tumor immune response, and is at least 10-50% above the basal (*i.e.*, untreated) level. Such response can be monitored by measuring the anti-tumor antibodies in a patient or by vaccine-dependent generation of cytolytic effector cells capable of killing the patient's tumor cells *in vitro*. Such vaccines should also be capable of causing an immune response that leads to an improved clinical outcome (*e.g.*, more

frequent remissions, complete or partial or longer disease-free survival) in vaccinated patients as compared to non-vaccinated patients. In general, for pharmaceutical compositions and vaccines comprising one or more polypeptides, the amount of each polypeptide present in a dose ranges from about 25 μg to 5 mg per kg of host. Suitable dose sizes will vary with the size of the patient, but will typically range from about 0.1 mL to about 5 mL.

In general, an appropriate dosage and treatment regimen provides the active compound(s) in an amount sufficient to provide therapeutic and/or prophylactic benefit. Such a response can be monitored by establishing an improved clinical outcome (*e.g.*, more frequent remissions, complete or partial, or longer disease-free survival) in treated patients as compared to non-treated patients. Increases in preexisting immune responses to a tumor protein generally correlate with an improved clinical outcome. Such immune responses may generally be evaluated using standard proliferation, cytotoxicity or cytokine assays, which may be performed using samples obtained from a patient before and after treatment.

Cancer Detection and Diagnostic Compositions, Methods and Kits

In general, a cancer may be detected in a patient based on the presence of one or more CT or CP proteins and/or polynucleotides encoding such proteins in a biological sample (for example, blood, sera, sputum urine and/or tumor biopsies) obtained from the patient. In other words, such proteins may be used as markers to indicate the presence or absence of a cancer. In addition, such proteins may be useful for the detection of other cancers. The binding agents provided herein generally permit detection of the level of antigen that binds to the agent in the biological sample. Polynucleotide primers and probes may be used to detect the level of mRNA encoding a tumor protein, which is also indicative of the presence or absence of a cancer. In general, a CT or CP tumor sequence should be present at a level that is at least three fold higher in tumor tissue than in normal tissue

There are a variety of assay formats known to those of ordinary skill in the art for using a binding agent to detect polypeptide markers in a sample. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual,* Cold Spring Harbor Laboratory, 1988. In general, the presence or absence of a cancer in a patient may be determined by (a) contacting a biological sample obtained from a patient with a binding agent; (b) detecting in the sample a level of polypeptide that binds to the binding agent; and (c) comparing the level of polypeptide with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of binding agent immobilized on a solid support to bind to and remove the polypeptide from the remainder of the sample. The bound polypeptide may then be detected using a detection reagent that contains a reporter group and specifically binds to the binding agent/polypeptide complex. Such detection reagents may comprise, for example, a binding agent that specifically binds to the polypeptide or an antibody or other agent that specifically binds to the binding agent, such as an anti-immunoglobulin, protein G, protein A or a lectin. Alternatively, a competitive assay may be utilized, in which a polypeptide is labeled with a reporter group and allowed to bind to the immobilized binding agent after incubation of the binding agent with the sample. The extent to which components of the sample inhibit the binding of the labeled polypeptide to the binding agent is indicative of the reactivity of the sample with the immobilized binding agent. Suitable polypeptides for use within such assays include full length CT or CP tumor proteins and polypeptide portions thereof to which the binding agent binds, as described above.

The solid support may be any material known to those of ordinary skill in the art to which the tumor protein may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The binding agent may be immobilized on the solid support using a variety of techniques known to those of skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "immobilization" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the agent and functional groups on the support or may be a linkage by way of a cross-linking agent). Immobilization by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the binding agent, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and about 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of binding agent ranging from about 10 ng to about 10 μg, and preferably about 100 ng to about 1 μg, is sufficient to immobilize an adequate amount of binding agent.

Covalent attachment of binding agent to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that will react with both the support and a functional group, such as a hydroxyl or amino group, on the binding agent. For example, the binding agent may be covalently attached to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the binding partner (*see, e.g.*, Pierce Immunotechnology Catalog and Handbook, 1991, at A12-A13).

In certain embodiments, the assay is a two-antibody sandwich assay. This assay may be performed by first contacting an antibody that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that polypeptides within the sample are allowed to bind to the immobilized antibody. Unbound sample is then removed from the immobilized polypeptide-antibody complexes and a detection reagent (preferably a second antibody capable of binding to a different site on the polypeptide) containing a reporter group is added. The amount of detection

reagent that remains bound to the solid support is then determined using a method appropriate for the specific reporter group.

More specifically, once the antibody is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or Tween 20™ (Sigma Chemical Co., St. Louis, MO). The immobilized antibody is then incubated with the sample, and polypeptide is allowed to bind to the antibody. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of polypeptide within a sample obtained from an individual with cancer. Preferably, the contact time is sufficient to achieve a level of binding that is at least about 95% of that achieved at equilibrium between bound and unbound polypeptide. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% Tween 20™. The second antibody, which contains a reporter group, may then be added to the solid support. Preferred reporter groups include those groups recited above.

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound polypeptide. An appropriate amount of time may generally be determined by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally

appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of a cancer, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value for the detection of a cancer is the average mean signal obtained when the immobilized antibody is incubated with samples from patients without the cancer. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for the cancer. In an alternate preferred embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett et al., *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., 1985, p. 106-7. Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100%-specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate, or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for a cancer.

In a related embodiment, the assay is performed in a flow-through or strip test format, wherein the binding agent is immobilized on a membrane, such as nitrocellulose. In the flow-through test, polypeptides within the sample

bind to the immobilized binding agent as the sample passes through the membrane. A second, labeled binding agent then binds to the binding agent-polypeptide complex as a solution containing the second binding agent flows through the membrane. The detection of bound second binding agent may then be performed as described above. In the strip test format, one end of the membrane to which binding agent is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing second binding agent and to the area of immobilized binding agent. Concentration of second binding agent at the area of immobilized antibody indicates the presence of a cancer. Typically, the concentration of second binding agent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of binding agent immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of polypeptide that would be sufficient to generate a positive signal in the two-antibody sandwich assay, in the format discussed above. Preferred binding agents for use in such assays are antibodies and antigen-binding fragments thereof. Preferably, the amount of antibody immobilized on the membrane ranges from about 25 ng to about 1μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount of biological sample.

Of course, numerous other assay protocols exist that are suitable for use with the tumor proteins or binding agents of the present invention. The above descriptions are intended to be exemplary only. For example, it will be apparent to those of ordinary skill in the art that the above protocols may be readily modified to use tumor polypeptides to detect antibodies that bind to such polypeptides in a biological sample. The detection of such tumor protein specific antibodies may correlate with the presence of a cancer.

A cancer may also, or alternatively, be detected based on the presence of T cells that specifically react with a tumor protein in a biological sample. Within certain methods, a biological sample comprising $CD4^+$ and/or

$CD8^+$ T cells isolated from a patient is incubated with a tumor polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with polypeptide (*e.g.*, 5 - 25 μg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of tumor polypeptide to serve as a control. For $CD4^+$ T cells, activation is preferably detected by evaluating proliferation of the T cells. For $CD8^+$ T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a cancer in the patient.

As noted above, a cancer may also, or alternatively, be detected based on the level of mRNA encoding a tumor protein in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a polynucleotide encoding the tumor protein. The amplified cDNA is then separated and detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes that specifically hybridize to a polynucleotide encoding a tumor protein may be used in a hybridization assay to detect the presence of polynucleotide encoding the tumor protein in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a tumor protein of the invention that is at least 10 nucleotides, and preferably at least 20

nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence as disclosed herein. Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis et al., *Cold Spring Harbor Symp. Quant. Biol.*, *51*:263 (1987); Erlich ed., *PCR Technology*, Stockton Press, NY, 1989).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as biopsy tissue, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

In another embodiment, the compositions described herein may be used as markers for the progression of cancer. In this embodiment, assays as described above for the diagnosis of a cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year, and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not

progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

As noted above, to improve sensitivity, multiple tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds, reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

Alternatively, a kit may be designed to detect the level of mRNA encoding a tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a tumor protein. Such an

oligonucleotide may be used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a tumor protein.

The following Examples are offered by way of illustration and not by way of limitation.

EXAMPLES

EXAMPLE 1

ISOLATION OF cDNA SEQUENCES FROM TESTIS AND PLACENTA LIBRARIES

This Example illustrates the identification of cDNA molecules encoding CT or CP tumor proteins.

To identify tumor antigens, a cDNA library was constructed from pooled placenta and testis mRNA. cDNA sequences isolated from the these subtracted libraries were searched against public databases including Genbank and included in Table 2 (SEQ ID NO:1-14). Two subtractive cDNA libraries (S1 and S2) were generated using placenta and testis as the testers and cDNA from normal tissue including lung, esophagus, liver, heart, skin, PBMC, brain, kidney, and pancreas as the driver. All libraries were generated by conventional biotin-streptavidin subtraction. The testers were cut with *Not I* and *SpeI* while all drivers were cut with *Bam*HI and *XhoI*. Hybridization of tester and driver was performed at 68°C overnight and followed by two successive biotin-strepavidin subtractions. The remaining cDNA was ligated into pBCSK+ plasmid vector. Sequences isolated using this method represent attractive targets for both diagnostic and immunotherapeutic applications.

cDNA sequences isolated from the first subtracted library S1 were searched against public databases including Genbank and included in Table 3 (SEQ ID NO: 15-21). This first subtractive library was found to contain a high percentage of a placenta specific protein, PL-4, therefore a second subtracted cDNA library, S2, was generated by including cDNA from PL-4 with the driver cDNA. cDNA sequences from this subtracted library were searched against

public databases including Genbank (SEQ ID NO:22-193). Those sequences showing some degree of similarity to known sequences in Genbank and those sequences that showed no significant similarity to any known sequences were identified and listed in Table 4.

Table 2

Summary of Placental and Testis cDNA Library

SEQ ID NO:	CLONE ID	GenBank
1	51572	h. thyroid hormone receptor-associated protein complex component TRAP100, also known as KIAA0130
2	51573	h. PL-4 hormone
3	51578	Novel, DKFZp568E1519
4	51581	TIAM2, T-cell Lymphoma invasion & mets 2
5	51582	SERP1, similar to rat ribosomal attached membrane prot. 4 (RAMP4)
6	51583	secreted SPARC/osteonectin
7	51584	Novel
8	51585	similar to rat Syntaxin 5
9	51586	Novel, KIAA 1084, chromosomal 19
10	51587	Novel, chromosomal 16
11	51590	Novel, chromosomal 17 DKFXp564c1563
12	51591	Novel
13	51592	CRES
14	51594	HCF1

Table 3

Summary of Placental and Testis Subtracted cDNA Library S1

SEQ ID NO:	CLONE ID	GenBank
15	54640	h. WS basic-helix-loop-helix eucine zipper protein
16	54652	h. ret finger protein 2 from CLL
17	54658	Novel, clone NT2RP4000498 seq from teratocarcinoma
18	54668	Novel
19	54673	Novel, h. PRO0529 mRNA from fetal liver
20	54675	Novel, h. unknown mRNA from acute promyelocytic Leuk
21	54676	h. malignant melanoma metastasis-suppressor (KISS-1) gene

Table 4

Summary of Placental and Testis Subracted cDNA Library S2

SEQ ID NO:	CLONE ID	GenBank
22	55217	Novel, KIAA0396
23	55218	stromrylsin-3
24	55220	Novel
25	55221	mammaglobin B precursor
26	55225	phosphotase methylesterase-1 PME-1
27	55227	Novel, BAC clone chro. 7
28	55228	Novel
29	55231	KIAA1046, T-cell specific Tyr kinase
30	55238	Novel
31	55241	Novel, BAC clone chrom. 2
32	55243	beta-cop homolog
33	55244	BTF3
34	55247	Novel
35	55249	Novel
36	55250	Novel
37	55252	Novel
38	55254	FGF receptor-1
39	55256	unknown mRNA
40	55257	Novel
41	55258	CAG repeat protein
42	55260	Novel, cDNA clone FLJ10983
43	55261	beta-adaptin
44	55262	Zinc finger TF
45	56372	Novel
46	56374	Novel
47	56375	Novel
48	56376	Novel
49	56377	Novel
50	56380	3-beta hydroxysteriod dehydrogenase
51	56382	MAGE XP-2
52	56387	Novel
53	56390	Novel
54	56391	Novel
55	56392	PLAB, TGF-beta family
56	56393	Novel
57	56394	Stromylysin-3
58	56395	Novel
59	56396	Novel
60	56397	Novel
61	56398	Novel

SEQ ID NO:	CLONE ID	GenBank
62	56399	MCS, mitochondrial capsule seleno protein
63	45401	3-beta hydroxylsteriod dehydrogenase isomerase
64	56402	Novel
65	56410	Novel
66	56413	Novel
67	56415	ASH1
68	56416	Novel
69	56418	Novel
70	56419	GTPase beta-sub. Homolog.
71	56421	Novel
72	56423	Novel
73	56424	GTP binding protein, development. regulated
74	56428	Novel
75	56429	SCR3
76	56432	Novel, KIAA0667
77	56433	UBP
78	56437	TH1 or HSPC
79	56439	Novel, KIAA1243
80	56441	CDR2
81	56445	Novel, KIAA0300
82	56448	MBP
83	56454	BPGM
84	56455	proteasome subunit, HC2
85	56457	Novel
86	56460	Enzyme
87	56462	CAGH3
88	56463	Novel
89	56467	VRK1
90	56965	Novel
91	56970	Novel, chromosomal 5
92	56971	EFG receptor precursor
93	56972	Novel
94	56974	Novel, chromosomal 9
95	56976	Novel
96	56979	Kiss-1
97	56980	??
98	56985	Novel
99	56989	PSG4
100	56990	Novel
101	56994	Novel
102	56995	Novel
103	56997	Novel, rat zinc finger protein
104	56998	HERQ
105	57004	Novel
106	57006	Novel

SEQ ID NO:	CLONE ID	GenBank
107	57008	Novel
108	57009	Novel
109	57011	DPH2L
110	57012	Novel
111	57015	angiotestin II receptor
112	57017	Novel, chromosomal 1
113	57020	GP-83
114	57025	Novel
115	57027	KBF2
116	57028	Novel
117	57032	Novel
118	57033	EF-1B
119	57034	h3.3
120	57041	Novel
121	57045	Novel
122	57050	rab geranylgeranyl transferase
123	57052	USP8, KIAA 0055
124	57055	Novel, KIAA1343
125	57057	Novel, chromosomal 1q12
126	57058	Novel
127	57472	h. YEAF1
128	57481	Novel, cDNA DKFZp564D0462
129	57483	Novel
130	57485	Novel, chromosomal 17
131	57486	Novel
132	57487	Novel
133	57489	Novel
134	57490	Novel
135	57491	Novel, cDNA DKFZp566B0846 homolog to mouse nectin-3
136	57512	TF AIB3
137	57517	PGDH
138	57528	90% to many chrom. Seq
139	57532	h. DBF4-like protein
140	57533	Novel
141	57534	BPGM enzyme
142	57540	Novel, cDNA FLJ10005 fis
143	57543	Novel
144	57545	h cDNA FLJ10944 fis candidate tumor suppressor for B-CLL
145	57546	H Hic-5
146	57548	Novel
147	57561	Novel
148	57563	Novel, cDNA FLJ20156
149	57565	Novel, chrom. 21q21.1-q21.2

SEQ ID NO:	CLONE ID	GenBank
150	57919	h melanocyte-specific gene (msg1)
151	57921	Novel
152	57922	Novel
153	57923	Novel
154	57933	KIAA0287, ch19
155	57934	h sin3 assos. Peptide SAP18
156	57937	Novel
157	57939	HSP89-alpha
158	57941	Pag
159	57945	Novel
160	57946	ch6 sequence
161	57948	h mRNA for GCM motif prot.
162	57949	h. pericentriol material-1 auto ag
163	57951	chr 7 seq
164	57952	CDNA FLJ10761 Fis
165	57953	transition protein TP-1
166	57956	Novel
167	57958	prostaglandin D synthase
168	57959	KIAA0617
169	57961	Novel
170	57965	Novel
171	57966	Novel
172	57971	KIAA0635
173	57976	inhibitor 2 of phosphatase 1
174	57977	procollagen C-proteinase pCP-2
175	57983	novel, chr 5 seq
176	57984	chronic gonadotropin HCG gene 5
177	57986	beta adaptin
178	57988	Novel
179	57991	PLU-1
180	57992	Novel
181	57998	pregnancy specific beta-1 glycoprotein
182	58000	dlk, putative homeotic Protein
183	58002	Novel, chr 21
184	58007	Novel, chr 7
185	58008	ran binding protein 2, membrane
186	58009	Novel, DKFZp564F053
187	59010	PEA 15
188	58013	Novel
189	58014	meltrin-S (ADAM12)
190	58015	Novel, chr 1 seq
191	58019	Novel
192	58025	h vanilloid receptor gene
193	58029	chr 14 seq

EXAMPLE 2

ISOLATION OF ADDITIONAL cDNA SEQUENCES FROM TESTIS AND PLACENTA LIBRARIES

332 cDNA sequences obtained from the S2 subtractive library described in Example 1 were searched against a public databases including Genbank. Some of those sequences showed some degree of similarity to known sequences in Genbank and some cDNA sequences showed no significant similarity to any known sequences. These sequences are disclosed in SEQ ID NO: 194-525 (see Table 5 for details).

Table 5

SEQ ID NO	Clone ID
194	62112394
195	62112395
196	62112396
197	62112399
198	62112403
199	62112404
200	62112405
201	62112406
202	62112412
203	62112416
204	62112419
205	62112421
206	62112423
207	62112428
208	62112429
209	62112431
210	62112437
211	62112439
212	62112440

SEQ ID NO	Clone ID
213	62112442
214	62112443
215	62112444
216	62112448
217	62112451
218	62112455
219	62112459
220	62112479
221	62112485
222	62116752
223	62116753
224	62116754
225	62116755
226	62116760
227	62116762
228	62116765
229	62116766
230	62116770
231	62116772
232	62116775
233	62116777
234	62116778
235	62116779
236	62116781
237	62116785
238	62116786
239	62116791
240	62116794
241	62116795
242	62116796
243	62116797
244	62116798

SEQ ID NO	Clone ID
245	62116799
246	62116801
247	62116803
248	62116804
249	62116805
250	62116807
251	62116809
252	62116810
253	62116824
254	62116825
255	62116826
256	62116828
257	62116831
258	62116832
259	62116836
260	62116837
261	62116838
262	62116839
263	62116841
264	62116842
265	61674531
266	61674535
267	61674537
268	61674538
269	61674539
270	61674540
271	61674547
272	61674548
273	61674551
274	61674552
275	61674555
276	61674558

SEQ ID NO	Clone ID
277	61674559
278	61674561
279	61674565
280	61674567
281	61674571
282	61674572
283	61674575
284	61674578
285	61674580
286	61674582
287	61674583
288	61674584
289	61674585
290	61674586
291	61674587
292	61674589
293	61674592
294	61674594
295	61674595
296	61674604
297	61674605
298	61674607
299	61674621
300	61674622
301	61674159
302	61674160
303	61674162
304	61674165
305	61674168
306	61674170
307	61674172
308	61674173

SEQ ID NO	Clone ID
309	61674174
310	61674177
311	61674178
312	61674179
313	61674180
314	61674184
315	61674185
316	61674186
317	61674187
318	61674188
319	61674189
320	61674190
321	61674194
322	61674196
323	61674198
324	61674199
325	61674201
326	61674202
327	61674203
328	61674204
329	61674205
330	61674206
331	61674207
332	61674211
333	61674212
334	61674213
335	61674215
336	61674217
337	61674218
338	61674219
339	61674220
340	61674221

SEQ ID NO	Clone ID
341	61674223
342	61674226
343	61674229
344	61674230
345	61674231
346	61674232
347	61674233
348	61674234
349	61674236
350	61674238
351	61674240
352	61674241
353	61674247
354	61674249
355	61674251
356	62116008
357	62116009
358	62116010
359	62116012
360	62116014
361	62116023
362	62116027
363	62116035
364	62116036
365	62116044
366	62116048
367	62116060
368	62116065
369	62116072
370	62116075
371	62116080
372	62116084

SEQ ID NO	Clone ID
373	62116089
374	62116092
375	62116096
376	62489678
377	62489685
378	62489686
379	62489687
380	62489688
381	62489692
382	62489693
383	62489694
384	62489695
385	62489696
386	62489698
387	62489699
388	62489700
389	62489701
390	62489702
391	62489703
392	62489704
393	62489705
394	62489706
394	62489710
396	62489711
397	62489712
398	62489713
399	62489714
400	62489716
401	62489717
402	62489718
403	62489719
404	62489720

SEQ ID NO	Clone ID
405	62489721
406	62489722
407	62489723
408	62489724
409	62489725
410	62489727
411	62489729
412	62489730
413	62489732
414	62489733
415	62489734
416	62489735
417	62489736
418	62489739
419	62489741
420	62489742
421	62489744
422	62489745
423	62489750
424	62489751
425	62489753
426	62489754
427	62489755
428	62489756
429	62489757
430	62489758
431	62489762
432	62489763
433	62489764
434	62489766
435	62489767
436	62489768

SEQ ID NO	Clone ID
437	62489769
438	62489770
439	62481819
440	62481820
441	62481821
442	62481822
443	62481823
444	62481824
445	62481825
446	62481826
447	62481827
448	62481829
449	62481830
450	62481831
451	62481832
452	62481834
453	62481835
454	62481836
455	62481837
456	62481838
457	62481839
458	62481840
459	62481843
460	62481844
461	62481845
462	62481846
463	62481847
464	62481848
465	62481849
466	62481850
467	62481851
468	62481853

SEQ ID NO	Clone ID
469	62481854
470	62481860
471	62481862
472	62481863
473	62481867
474	62481869
475	62481870
476	62481880
477	62481881
478	62481883
479	62481884
480	62481887
481	62481890
482	62481894
483	62481900
484	62481903
485	62481904
486	62481905
487	62481907
488	62481909
489	62481910
490	62416698
491	62416700
492	62416701
493	62416702
494	62416703
495	62416705
496	62416707
497	62416708
498	62416710
499	62416713
500	62416714

SEQ ID NO	Clone ID
501	62416715
502	62416716
503	62416719
504	62416723
505	62416724
506	62416726
507	62416727
508	62416728
509	62416730
510	62416731
511	62416733
512	62416734
513	62416738
514	62416740
515	62416741
516	62416742
517	62416748
518	62416749
519	62416752
520	62416756
521	62416764
522	62416769
523	62416770
524	62416771
525	62416772

EXAMPLE 3

ISOLATION OF ADDITIONAL cDNA SEQUENCES FROM TESTIS AND PLACENTA LIBRARIES

It has been well documented that tumor specific antigens are not only expressed in a wide variety of tumors, but also in placenta and testis.

Here, sequence data has been analyzed for differential gene expression utilizing an approach that is referred to as electronic subtraction. Specifically, a large database of human EST sequences was analyzed using sequence similarity and clustering algorithms to identify genes that may show differential expression in tumors as well as placenta and testis. Here sequences of EST clones were derived from 36 libraries, including cancers of germ cells and testis as well as normal placenta and testis. The subtraction was generated using 212,949 tester cDNAs generated from the above libraries and 439,248 driver cDNAs derived from normal tissue libraries. The subtraction criteria were set so that only tester cDNAs that had no matches in the normal tissue cDNAs were scored. This resulted in the identification of 36,386 cDNAs. After eliminating redundant clones, the number of cDNAs of interest was reduced to 32,969. These sequences were then divided into three groups. Group I consists of sequences which were only represented once (17,256) cDNAs, Group II consists of clones which were represented two or three times (8717) cDNAs and Group III contains cDNAs represented four or more times (6996) cDNAs.

Group III cDNAs were then assessed for the relative abundance in selected cDNA libraries, such as germ cell tumor library and testis tumor library, using Transcript Imaging Analysis. In order for a sequence to be selected, ≥70% of the hits obtained must be from either the germ cell tumor library or the testis tumor library. Sequences isolated using this method represent attractive targets for both diagnostic and immunotherapeutic applications. 45 cDNAs showed good electronic expression and are disclosed in SEQ ID NOs:526-570. 19 of these cDNAs were PCR amplified from testis cDNA and confirmed by sequence analysis.

EXAMPLE 4

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY

In additional studies, sequences disclosed herein are evaluated for overexpression in specific tumor tissues by microarray analysis. Using this

approach, cDNA sequences are PCR amplified and their mRNA expression profiles in tumor and normal tissues are examined using cDNA microarray technology essentially as described (Shena, M. et al., 1995 Science *270*:467-70). In brief, the clones are arrayed onto glass slides as multiple replicas, with each location corresponding to a unique cDNA clone (as many as 5500 clones can be arrayed on a single slide, or chip). Each chip is hybridized with a pair of cDNA probes that are fluorescence-labeled with Cy3 and Cy5, respectively. Typically, 1μg of polyA$^+$ RNA is used to generate each cDNA probe. After hybridization, the chips are scanned and the fluorescence intensity recorded for both Cy3 and Cy5 channels. There are multiple built-in quality control steps. First, the probe quality is monitored using a panel of ubiquitously expressed genes. Secondly, the control plate also can include yeast DNA fragments of which complementary RNA may be spiked into the probe synthesis for measuring the quality of the probe and the sensitivity of the analysis. Currently, the technology offers a sensitivity of 1 in 100,000 copies of mRNA. Finally, the reproducibility of this technology can be ensured by including duplicated control cDNA elements at different locations.

EXAMPLE 5

ANALYSIS OF cDNA EXPRESSION USING MICROARRAY TECHNOLOGY AND REAL-TIME PCR

cDNA clones from the S1 and S2 subtraction libraries generated in Example 1 were attached to a microarray chip to determine their expression profiles by microarray analysis. The libraries were screened using probes designed from a variety of tumor and normal tissues. Expression analysis demonstrated that a number of clones were overexpressed in tumor samples. Several of these clones were then selected for further expression analysis using Real-time PCR.

For example, clone 56392 (SEQ ID NO:55), a TGF-β superfamily protein member, was over expressed in lung adenocarcinoma and lung pleural infusions but was not detected in normal lung tissue. This clone was also

detected in normal pancreas tissue. Clone 56396 (SEQ ID NO:59) was overexpressed in lung squamous, with significantly lower levels of expression in normal tissue, including normal lung. Clone 56372 (SEQ ID NO:45), a novel protein, was over expressed in atypical carcinoid metastasis, with little or no expression in normal tissues. Clone 57490 (SEQ ID NO:134) was over expressed in approximately 75% of lung tumors and lung tumor cell lines tested and undetectable in normal lung tissue. . Some expression was also significant in normal brain, pituitary gland, and trachea. These expression profiles indicate that the sequences will have utility, for example in the detection of lung cancers

EXAMPLE 6

ISOLATION OF ADDITIONAL cDNA SEQUENCES FROM TESTIS AND PLACENTA LIBRARIES

Example 3 discusses the generation of an electronic subtraction library using placenta and testis as drivers. From this library, 6996 Group III cDNAs were selected. These cDNAs were represented four or more times in tumor specific libraries. The clones were subjected to electronic Northern analysis and were prioritized according to the ratio of EST sequences present in all tumor libraries compared to the total number of EST hits. Disclosed as SEQ ID NOs:571-912 are the cDNA sequences for clones which had a ratio number of no less than 0.4 as shown in Table 6.

Table 6

cDNA sequences for Group III Clones With Good Expression Profiles in Tumors

SEQ ID NO.	CLONE ID	Genbank
571	079399.1	
572	446653.1	
573	431774.1	
574	454039.1	
575	898572.3	AAG30289.1 g11096305 NALP2 0
576	986945.1	
577	182036.1	

SEQ ID NO.	CLONE ID	Genbank
578	392306.1	
579	392759.1	
580	394030.1	
581	433575.1	
582	437485.1	
583	477078.1	
584	986501.1	
585	987387.1	
586	065609.1	AJ289880 g8980366 Human KIAA0851 gene (partial), XT3 gene and LZTFL1 gene. 0
587	392088.1	
588	392333.1	
589	415611.1	
590	987619.1	
591	045858.1	
592	309422.1	
593	316712.1	
594	391944.1	
595	432789.1	
596	448143.1	
597	985976.1	
598	987809.1	
599	239505.1	
600	258518.1	U29112 g940057 Human lipoma cell line Li-538/SV40 ectopic sequence from HMGI-C fusion mRNA, 3' sequence. 0
601	305588.1	AL163243 g7768708 Human genomic DNA, chromosome 21q, section 42/105. 1e-08
602	308950.1	Z60140 g1032053 Human CpG island DNA genomic Mse1 fragment, clone 190c3, reverse read cpg190c3.rt1a. 0
603	326094.1	AF084198 g6682357 Orangutan (Pongo pygmaeus) gamma-aminobutyric acid receptor A5 subunit duplicated gene, 5'UTR region. 1e-26
604	339329.1	
605	394029.1	
606	405890.1	AK000310 g7020308 Human cDNA FLJ20303 fis, clone HEP06676. 1e-13
607	432323.1	AAA40905.1 g203396 potassium channel 6e-76
608	433390.1	
609	437615.1	U76666 g1809330 Human N-type calcium channel alpha1B subunit gene, 5' upstream region and partial cds. 6e-33
610	442116.1	
611	446955.1	

SEQ ID NO.	CLONE ID	Genbank
612	450025.1	
613	462504.1	
614	464221.1	
615	960143.1	
616	986197.1	
617	986399.1	
618	986489.1	AL022729 g3093310 Human gene similar to Z.mays ras-like (X63277) and Human RAY1 (X79781). 3e-17
619	986751.1	
620	986787.1	
621	987125.1	
622	987521.1	
623	987595.1	
624	1043767.1	
625	1066489.1	
626	1098640.1	AF113697 g6855629 Human clone FLB5332. 0
627	402707.1	
628	402419.1	AK025715 g10438324 Human cDNA: FLJ22062 fis, clone HEP10276, highly similar to AF151890 Human CGI-132 protein mRNA. 2e-13
629	172918.1	
630	448803.1	
631	343260.1	
632	394027.1	AF240580 g9408534 Human clone 17ptel_2111ctg_drft sequence. 2e-09
633	987177.1	
634	475616.1	J02986 g184430 Human transforming protein (hst) gene, complete cds. 0
635	171127.3	
636	399581.1	AB019439 g4512287 Human DNA for Ig heavy-chain variable region, complete sequence, 3 of 5. 0
637	421371.1	AL163205 g7768749 Human genomic DNA, chromosome 21q, section 4/105. 2e-76
638	440465.1	AF187320 g6164847 Human transferrin receptor (TFRC) gene, complete cds. 0
639	127864.1	
640	017185.1	
641	047260.1	AP000355 g5103018 Human genomic DNA, chromosome 22q11.2, clone KB1896H10. 3e-11
642	206513.1	
643	312192.1	AL163259 g7768677 Human genomic DNA, chromosome 21q, section 58/105. 0
644	326968.1	
645	343222.1	

SEQ ID NO.	CLONE ID	Genbank
646	392400.1	
647	416535.1	
648	455418.1	
649	985965.1	
650	986659.1	
651	336127.1	
652	432083.1	
653	021641.1	AL163204 g7717247 Human chromosome 21 segment HS21C004. 7e-16
654	317808.1	
655	316185.1	
656	336811.1	CAB55923.1 g5911875 hypothetical protein 4e-18
657	047316.1	
658	066142.1	
659	143226.1	
660	181356.1	
661	199943.1	
662	219193.1	
663	311931.1	
664	360081.3	
665	394002.1	
666	396314.1	
667	402516.1	
668	410836.1	AL163240 g7768704 Human genomic DNA, chromosome 21q, section 39/105. 5e-18
669	413883.1	AE000521 g2358019 Human T-cell receptor alpha delta locus from bases 1 to 250529 (section 1 of 5) of the Complete Nucleotide Sequence. 1e-34
670	415477.1	
671	431905.1	
672	432771.1	U82670 g7274891 Human chromosome Xq28 psHMG17 pseudogene, complete sequence; and melanoma antigen family A1 (MAGEA1) and zinc finger protein 275 (ZNF275) genes, complete cds. 6e-15
673	434300.1	
674	439683.1	
675	447331.1	
676	447353.1	
677	448396.1	
678	450955.1	
679	955279.1	
680	958598.1	
681	985594.1	
682	985878.1	

SEQ ID NO.	CLONE ID	Genbank
683	986790.1	
684	986967.1	
685	987378.1	
686	987914.1	
687	987959.1	
688	404898.1	AL163301 g7717439 Human chromosome 21 segment HS21C101. 0
689	986832.1	
690	194683.1	
691	386470.1	
692	438122.1	AF240580 g9408534 Human clone 17ptel_2111ctg_drft sequence. 0
693	455419.1	
694	891444.3	
695	986202.1	
696	312226.1	AK026368 g10439211 Human cDNA: FLJ22715 fis, clone HSI13726. 3e-97
697	392760.1	
698	107221.1	
699	142773.1	
700	392101.1	
701	002431.1	
702	058453.1	
703	108294.1	
704	163554.1	
705	170736.1	
706	224797.1	
707	226946.1	
708	231098.1	
709	239165.1	
710	309987.1	AL163210 g7717240 Human chromosome 21 segment SH21C001. 0
711	311358.1	
712	312221.1	
713	314029.1	
714	316239.1	
715	321442.1	Z54527 g1020568 Human CpG island DNA genomic Mse 1 fragment, clone 12c4, reverse read cpg12c4.451d. 1e-73
716	341275.1	
717	343075.1	
718	386344.1	
719	392345.1	
720	399840.1	

SEQ ID NO.	CLONE ID	Genbank
721	400295.1	AB031325 g7023984 Human gene for calcium-sensing receptor, exons, promotor region. 0
722	402116.1	
723	402496.1	
724	416014.1	
725	432129.1	AB009460 g2924559 Human klotho gene for Klotho protein (secreted form and membrane form), exon 2, exon 3, exon 4, exon 5 and complete cds. 7e-29
726	436520.1	
727	443583.1	AJ289880 g8980366 Human KIAA0851 gene (partial), XT3 gene and LZTFL1 gene. 0
728	448587.1	
729	456183.1	
730	475859.1	AAC72810.1 g3599347 ORF2 5e-08
731	957898.1	AAC51273.1 g2072967 putative p 150 0.0002
732	986470.1	
733	986625.1	
734	987215.1	
735	987277.1	
736	987693.1	AE000521 g2358019 Human T-cell receptor alpha delta locus from bases 1 to 250529 (section 1 of 5) of the Complete Nucleotide Sequences. 3e-18
737	1031675.1	
738	1400146.1	AK027134 g10440186 Human cDNA: FLJ23481 fis, clone KAIA03003. 3e-78
739	043644.1	AK001746 g7023201 Human cDNA FLJ10884 fis, clone NT2RP4001950. 0
740	217013.1	
741	1076539.4	AK023721 g10435737 Human cDNA FLJ13659 fis, clone PLACE1011576, moderately similar to Human Kruppel related zinc finger protein (HTF10) mRNA. 8e-20
742	334794.1	
743	349909.1	S65622 g433902 Human mRNA for zinc finger protein. 0
744	027992.1	
745	067855.1	X02654 g37556 Human U4 small nuclear RNA pseudogene (U4/8). 4e-10
746	110476.1	
747	309032.1	
748	314806.3	AK001558 g7022883 Human cDNA FLJ10696 fis, clone NT24P3000484. 3e-47
749	318047.1	
750	386652.1	
751	403686.1	

SEQ ID NO.	CLONE ID	Genbank
752	410816.1	
753	430948.1	CAA59168.1 g673456 testis nuclear RNA binding protein 6e-65
754	434816.1	
755	435324.1	
756	476506.1	AF246983 g11228667 Human GNAS1 gene, exon 1A, complete sequence, alternatively spliced. 0
757	986243.1	
758	986474.1	
759	212176.1	AJ132429 g4775352 Human mRNA for hyperpolarization-activated cyclic nucleotide gated cation channel hHCN4. 0
760	251447.4	AJ278111 g8216986 Human mRNA for putative tumor antigen (SAGE gene). 0
761	334006.1	AJ001699 g2558580 Human mRNA for Brachyury (T) protein. 0
762	039276.1	
763	018023.1	AK025600 g10438168 Human cDNA FLJ21947 fis, clone HEP04896. 6e-19
764	406014.1	
765	032082.1	
766	001291.1	AF063012 g4377856 Human basic helix-loop-helix protein HAND1 (HAND1) gene, complete cds. 0
767	048219.1	
768	059333.1	AK023189 g10435003 Human cDNA FLJ13127 fis, clone NT2RP3002911. 0
769	179346.1	
770	219197.1	
771	324793.1	
772	401479.1	
773	985535.1	
774	144351.5	AK022643 g10434154 Human cDNA FLJ12581 fis, clone NT2RM4001140, weakly similar to HOMEOBOX PROTEIN MSH-D. 0
775	044189.1	
776	305582.2	
777	012033.1	
778	026761.1	
779	028035.1	
780	068777.1	
781	310200.1	
782	310775.1	
783	342490.2	
784	395607.1	
785	402228.1	

SEQ ID NO.	CLONE ID	Genbank
786	333170.7	BAA12350.1 g1304193 HMG2 1e-33
787	207494.1	AK002193 g7023919 Human cDNA FLJ11331 fis, clone PLACE1010547. 0
788	010614.1	
789	317847.1	
790	006860.1	
791	017164.1	
792	017166.1	
793	022907.1	L14561 g4165324 Human plasma membrane calcium ATPase isoform 1 (ATP2B1) gene, alternative splice products, partial cds. 0
794	046902.1	
795	051551.1	
796	064141.1	
797	083061.1	
798	107008.1	
799	153177.1	
800	208852.1	
801	209720.1	
802	219169.1	
803	223855.1	
804	229306.1	
805	332798.1	
806	412815.2	
807	446189.1	
808	958050.1	
809	960010.3	BAA28284.1 g3142153 Arx homeoprotein 1e-37
810	986400.1	Novel
811	987727.1	CAA71005.1 g1835124 unnamed protein product 4e-57
812	1383998.1	AF187850 g6166440 Human PPAR delta gene, promoter region. 3e-22
813	1387601.1	
814	008076.1	
815	028982.1	AF085831 g3483145 Human full length insert cDNA clone YI40A07. 0
816	230688.2	
817	018752.4	AF015950 g2330016 Human telomerase reverse transcriptase (hTRT) mRNA, complete cds. 0
818	210675.1	AK021465 g10432657 Human cDNA FLJ11403 fis, clone HEMBA10000726. 5e-80
819	351238.1	AE000521 g2358042 Human T-cell receptor alpha delta locus from bases 501613 to 752736 (section 3 of 5) of the Complete Nucleotide Sequence. 1e-41

SEQ ID NO.	CLONE ID	Genbank
820	230823.1	AK001666 g7023063 Human cDNA FLJ10804 fis, clone NT2RP4000837, weakly similar to Human mRNA for zinc finger protein SALL1. 0
821	306933.1	
822	017249.2	
823	033237.2	
824	070785.1	AK023843 g10435902 Human cDNA FLJ13781 fis, clone PLACE4000465. 0
825	210391.1	AF082338 g3435280 Human p14ARF gene, promotor region, complete sequence; and exon 1 beta, partial sequence. 0
826	308370.1	
827	315959.2	
828	327347.1	
829	345528.1	
830	361450.1	
831	395814.2	
832	402258.1	
833	405213.1	
834	431597.2	
835	986293.1	D80008 g1136431 Human mRNA for KIAA0186 gene, complete cds. 6e-89
836	1004207.1	AK023967 g10436117 Human cDNA FLJ13905 fis, clone THYRO1001907. 0
837	084452.1	
838	180755.1	AF111168 g5468517 Human serine palmitoyl transferase, subunit II gene, complete cds; and unknown genes. 0
839	229743.3	
840	206873.1	
841	438711.1	CAB87658.1 g7573352 MAP3K delta-1 protein kinase 9e-11
842	467434.1	Novel
843	349411.2	U10688 g533516 Human MAGE-4b antigen (MAGE4b) gene, complete cds. 0
844	014640.1	
845	123840.1	
846	179576.1	AAB34947.1 g1042009 genomic screen homeobox protein 2 [mice, NIH Swiss and ICR/Swiss-Webster, embryos, Peptide, 305 aa] 6e-47
847	256819.1	AF130063 g11493432 Human clone FLB7723 PRO2055 mRNA, complete cds. 2e-17
848	402568.1	
849	430738.3	AAC62527.1 g2581864 submaxillary apomucin 2e-55

SEQ ID NO.	CLONE ID	Genbank
850	432567.1	AK026061 g10438784 Human cDNA: FLJ22408 fis, clone HRC08416. 0
851	199655.3	U82811 g2662410 Human homeodomain-containing protein (HANF) mRNA, complete cds. 0
852	207286.1	AL163240 g7768704 Human genomic DNA, chromosome 21q, section 39/105. 0
853	254297.3	
854	1399268.2	D42044 g577300 Human mRNA for KIAA0090 gene, partial cds. 2e-79
855	210362.1	
856	183331.1	AF279289 g8927580 Human testes-specific heterogenous nuclear ribonucleoprotein G-T gene, partial cds. 0
857	336866.1	
858	346005.1	
859	414385.3	
860	981478.2	
861	180651.1	AAB97362.1 g2795887 unknown 2e-26
862	189061.3	
863	205792.1	AF162278 g6002918 Human Skn-1a/Epoc-1/Oct-11 POU transcription factor mRNA, complete cds. 0
864	342238.1	
865	416035.1	
866	479479.2	
867	034487.1	
868	156192.1	
869	161646.1	
870	214510.1	
871	257870.1	AF279780 g11095822 Human clone N11 Ntera2D1 teratocarcinoma mRNA
872	333776.4	
873	336561.1	
874	338031.1	
875	357004.3	AF090189 g4406153 Human Cerberus-related protein (CER1) gene, complete cds.
876	902956.2	
877	982913.2	
878	1040934.1	CAB88102.1 g7594599 dJ310013.4 (novel protein similar to predicted C. elegans an C. intesinalis proteins)
879	985115.2	
880	219737.8	AJ251026 g6900076 Human mRNA for putative odorant binding protein b-a (OBPIIb gene)
881	059654.1	AK001064 g7022100 Human cDNA FLJ10202 fis, clone HEMBA 1004929

SEQ ID NO.	CLONE ID	Genbank
882	346641.4	U20582 g684935 Human actin-like peptide mRNA, partial cds
883	982641.1	AP000501 g5926688 Huamn genomic DNA, Chrom 8p11.2, clone 91h23 to 9-41.
884	033028.1	
885	111612.1	
886	213081.1	
887	274780.1	AK023400 g1045322 Human cDNA FLJ13338 fis, clone OVARC1001883.
888	404725.1	AF305057 g11094017 Human RTS gene, complete cds, alternatively spliced
889	058790.1	
890	082509.1	
891	238883.2	
892	368015.2	AL163303 g7717449 human chrom 21, segment HS21C103
893	005223.1	
894	007157.1	
895	009880.1	
896	011701.1	
897	022523.1	
898	076749.1	
899	127051.1	
900	129948.1	
901	152856.1	
902	168601.1	
903	175164.1	
904	203627.1	
905	212044.1	
906	214286.1	
907	220557.1	
908	230796.1	AP001330 g8698845 Human genomic DNA, Chrom 8q23, clone KB1410C5
909	231673.1	
910	241027.1	AL163221 g7768692 Human genomic DNA Chrom. 21q, section 20/105
911	258898.1	
912	345082.1	

EXAMPLE 7

ISOLATION OF cDNA SEQUENCES OVER EXPRESSED IN FETAL TISSUE AND EMBRYO LIBRARIES

Previous examples have focused on the identification of tumor antigens that are over expressed in placenta and testis. It has been well documented that tumor specific antigens are over expressed not only in tumors, placenta and testis, but also in fetal tissues. These proteins are referred to as oncofetal proteins. This Example describes the use electronic subtraction to identify antigens that are over expressed in fetal tissue compared to normal adult tissue (this subtraction process is described in detail in Example 3).

The tester component of this subtraction was 27,955 sequences from 6 early fetal and embryo libraries with the drivers consisting of 723,535 sequences derived from 190 normal tissue libraries. The subtraction criterion was set so that only sequences that were unique to the tester libraries were retrieved. This process resulted in the identification 2874 sequences. These sequences were then submitted for electronic Northern (eNorthern) analysis, the process of which is described in more detail in Example 6. Sequences identified from the subtraction process were screened against two groups of libraries. Group I included 368 tumor tissue libraries and Group II included 370 libraries derived from normal tissues. After analysis, only 1058 sequences remained. These were divided into three groups. Group I include 49 sequences that had no hits in the normal tissue libraries and 2 hits in the tumor tissue libraries. Group II included 28 sequences that had one or more hits in the normal tissue libraries, but had a tumor:normal tissue ratio of ≥2. Group III included 31 sequences that demonstrated a tumor:normal tissue ratio of ≥2. The remaining sequences had either no hits in the tumor tissue libraries screened or had a tumor:normal tissue ratio of <2. The sequences from Groups I, II, and III and disclosed in SEQ ID Nos:913-1021 (see Table 7 for details).

Table 7

SEQ ID NO:	Clone ID	Description
913	205495	Novel
914	8613	Novel
915	16725	Novel
916	216611	HSA238982 HOMO SAPIENS MRNA FOR VNN3 PROTEIN
917	408621	XM_015373 Homo sapiens lens intrinsic membrane protein 2 (19kD) (LIM2)
918	14675	Novel
919	18911	Novel
920	28267	Novel
921	106832	AF229172 HOMO SAPIENS CLASS III MYOSIN (MYO3A)
922	181010	Novel
923	182020	Novel
924	264136	Novel
925	289838	S65125 UFO=PROTO-ONCOGENE
926	326211	Novel
927	336837	Novel
928	408388	HSHOX3D HUMAN HOX3D GENE FOR HOMEOPROTEIN HOX3D
929	439717	gi|10439535|dbj|AK026639.1|AK026639 Homo sapiens cDNA: FLJ22986 fis, clone KAT11742
930	8615	Novel
931	12285	Novel
932	22644	Novel
933	31287	HSU82970 HUMAN METALLOENDOPEPTIDASE HOMOLOG (PEX)
934	33837	Novel
935	41242	Novel
936	50247	AK021638 Homo sapiens cDNA FLJ11576 fis, clone HEMBA1003548
937	51790	L44140_HUMAN CHROMOSOME X
938	60218	Novel
939	61315	Novel
940	78499	Novel
941	147200	Novel
942	181754	Novel
943	182283	Novel
944	182701	Novel
945	183469	Splice-Form1_gi|514264 U07563 Human proto-oncogene tyrosine-protein kinase

SEQ ID NO:	Clone ID	Description
		(ABL)_Form1_gi|514264
946	183469	Splice-Form2_gi|514262 U07563 Human proto-oncogene tyrosine-protein kinase (ABL)_Form1_gi|514264
947	202850	Novel
948	219602	Novel
949	243305	AAC28425 gi|3378094 KRAB domain zinc finger protein (2e-46)
950	260832	Novel
951	268274	Novel
952	296634	Novel
953	305318	Novel
954	328415	Novel
955	336007	Novel
956	341359	Novel
957	386791	Novel
958	396684	Novel
959	403617	Novel
960	410378	HSGLYRA1 H.SAPIENS ALPHA-2 STRYCHNINE BINDING SUBUNIT OF INHIBITORY GLYCINE RECEPTOR MRNA
961	476586	Novel
962	903682	HUMCRYGQ6 HUMAN GAMMA-E-CRYSTALLIN PSEUDOGENE (GAMMA-G2-PSI), EXON 3
963	332511	Novel
964	14767	Novel
965	403717	Novel
966	336169	Novel
967	143671	Novel
968	29940	gi|348863|gb|L22396.1|HUM328MFD Homo sapiens DNA sequence, repeat region
969	35237	Novel
970	106651	Novel
971	106876	Novel
972	334655	Novel
973	30968	gi|1019788|emb|X89067.1|HSTRPC2GN H.sapiens mRNA for trpc2 transcript (possible pseudogene)
974	85246	AL163262 gi|7768678 Human genomic DNA, chromosome 21q, section 61/105. 0
975	219294	AF088219 gi|3719360 Human CC chemokine gene cluster, complete sequence. 0
976	403249	Novel
977	474581	KIAA0738 protein

SEQ ID NO:	Clone ID	Description
978	30863	Novel
979	181611	Novel
980	245112	Novel
981	336004	Novel
982	342160	Novel
983	415945	Novel
984	78626	gi|3901259|gb|AF053952.1|AF053952 Homo sapiens CBFA1/OSF2 transcription factor mRNA, partial cds
985	977607	Novel
986	7999	Novel
987	407637	gi|3142152|dbj|AB006103.1|AB006103 Mus musculus mRNA for Arx homeoprotein, complete cds
988	8560	Novel
989	403009	Novel
990	335824	gi|1841337|dbj|D82344.1|D82344 Homo sapiens mRNA for NBPhox, complete cds
991	180396	Novel
992	481845	AJ003147 g2808656 Human complete genomic sequence between D16S3070 and D16S3275, containing Familial Mediterranean Fever gene disease. 0
993	033024	Novel
994	179047	Novel
995	261144	Novel
996	264632	Novel
997	398220	Novel
998	403008	Novel
999	475943	Novel
1000	411221	Novel
1001	334008	Novel
1002	015564	AL163297 g7768757 Human genomic DNA, chromosome 21q, section 96/105. 0
1003	309028	Novel
1004	406447	Novel
1005	234435	Novel
1006	336327	Novel
1007	977839	Novel
1008	959783	gi|509242|emb|X73478|HSPTPAA H.sapiens hPTPA mRNA
1009	203703	Novel
1010	340914	Novel
1011	402141	gi|2588894|dbj|AB008515|AB008515 Homo sapiens mRNA for RanBPM, complete cds

SEQ ID NO:	Clone ID	Description
1012	211396	Novel
1013	336423	gi|10802800|gb|AF251187|AF251187 Homo sapiens periodontal ligament cell specific protein 2 mRNA, partial sequence
1014	206793	Novel
1015	334729	gi|488286|gb|U08336|HSU08336 Human basic helix-loop-helix transcription factor mRNA, complete cds
1016	442113	Novel
1017	183506	Novel
1018	213200	Novel
1019	182387	Novel
1020	197880	gi|6453515|emb|AL133073|HSM801344 Homo sapiens mRNA; cDNA DKFZp434L0217 (from clone DKFZp434L0217); partial cds
1021	343473	gi|9967304|dbj|AB047941|AB047941 Macaca fascicularis brain cDNA, clone:QnpA-11683

EXAMPLE 8

IDENTIFICATION OF cDNA SEQUENCES ENCODING TESTIS SPECIFIC ANTIGENS

This Example illustrates the identification of cDNA molecules encoding testis specific antigens that may be over expressed in a variety of tumor types. This example identifies the generation of the TPS1 library using a PCR based subtraction. The library was generated using testis cDNA as the tester, with the driver cDNA composed of eight normal tissues, including lung, trachea, liver, heart, bone marrow, brain, kidney, and pancreas.

The TPS1 cDNA library was constructed and cloned into the PCR2.1 vector (Invitrogen, Carlsbad, CA) by subtracting a pool of testis derived cDNAs with a pool of cDNA from normal tissues including lung, trachea, liver, heart, bone marrow, brain, kidney, and pancreas using PCR subtraction methodologies (Clontech, Palo Alto, CA). The subtraction was performed using a PCR-based protocol, which was modified to generate larger fragments. Within this protocol, tester and driver double stranded cDNA were separately digested with five restriction enzymes that recognize six-nucleotide restriction sites (MluI, MscI, PvuII, SalI and StuI). This digestion resulted in an average

cDNA size of 600 bp, rather than the average size of 300 bp that results from digestion with RsaI according to the Clontech protocol. This modification does not affect the subtraction efficiency. Two tester populations were then created with different adapters, and the driver library remained without adapters.

The tester and driver libraries were then hybridized using excess driver cDNA. In the first hybridization step, driver was separately hybridized with the tester cDNA population. This results in populations of (a) unhybridized tester cDNAs, (b) tester cDNAs hybridized to other tester cDNAs, (c) tester cDNAs hybridized to driver cDNAs, and (d) unhybridized driver cDNAs. The two separate hybridization reactions were then combined, and rehybridized in the presence of additional denatured driver cDNA. Following this second hybridization, in addition to populations (a) through (d), a fifth population (e) was generated in which tester cDNA with one adapter was hybridized to tester cDNA with the second adapter. Accordingly, the second hybridization step resulted in enrichment of differentially expressed sequences that can be used as templates for PCR amplification with adapter-specific primers. The ends were then filled in, and PCR amplification was performed using adapter-specific primers. Only population (e), which contained tester cDNA that did not hybridize to driver cDNA, was amplified exponentially. A second PCR amplification step was performed, to reduce background and further enrich differentially expressed sequences. This PCR-based subtraction technique normalizes differentially expressed cDNAs so that rare transcripts that are over-expressed in testis specific tissue may be recovered. Such transcripts would be difficult to recover by traditional subtraction methods.

Clones generated from the TPS1 library were sequenced to determine the identity of the inserts and these sequences are disclosed in SEQ ID NOs:1022-2522 (see Table 8 for details).

Table 8

SEQ ID NO:	Clone ID	Description
1022	68620	Novel
1023	68621	Novel

SEQ ID NO:	Clone ID	Description
1024	68622	Novel
1025	68623	Novel
1026	68624	Novel
1027	68625	Novel
1028	68626	Novel
1029	68628	Novel
1030	68629	Novel
1031	68630	Novel
1032	68631	Novel
1033	68633	Novel
1034	68634	Novel
1035	68636	Novel
1036	68637	Novel
1037	68638	Novel
1038	68639	Novel
1039	68640	Novel
1040	68641	Novel
1041	68644	Novel
1042	68645	Novel
1043	68647	Novel
1044	68648	Novel
1045	68649	Novel
1046	68650	Novel
1047	68651	Novel
1048	68652	Novel
1049	68653	Novel
1050	68654	Novel
1051	68655	Novel
1052	68656	Novel
1053	68657	Novel
1054	68658	Novel
1055	68659	Novel
1056	68661	Novel
1057	68662	Novel
1058	68663	Novel
1059	68664	Novel
1060	68667	Novel
1061	68668	Novel
1062	68669	Novel
1063	68670	Novel
1064	68671	Novel
1065	68673	Novel

SEQ ID NO:	Clone ID	Description
1066	68674	Novel
1067	68675	Novel
1068	68676	Novel
1069	68678	Novel
1070	68679	Novel
1071	68680	Novel
1072	68681	Novel
1073	68682	Novel
1074	68684	Novel
1075	68685	Novel
1076	68686	Novel
1077	68687	Novel
1078	68688	Novel
1079	68690	Novel
1080	68693	Novel
1081	68697	Novel
1082	68698	Novel
1083	68699	Novel
1084	68700	Novel
1085	68701	Novel
1086	68702	Novel
1087	68703	Novel
1088	68704	Novel
1089	68705	Novel
1090	68706	Novel
1091	68707	Novel
1092	68713	Novel
1093	64485262	R0760:A02
1094	64485264	R0760:A05
1095	64485265	R0760:A06
1096	64485266	R0760:A07
1097	64485267	R0760:A08
1098	64485268	R0760:A09
1099	64485269	R0760:A10
1100	64485270	R0760:A11
1101	64485271	R0760:A12
1102	64485272	R0760:B01
1103	64485273	R0760:B02
1104	64485274	R0760:B03
1105	64485275	R0760:B04
1106	64485276	R0760:B05
1107	64485277	R0760:B06

SEQ ID NO:	Clone ID	Description
1108	64485278	R0760:B07
1109	64485279	R0760:B08
1110	64485280	R0760:B09
1111	64485281	R0760:B10
1112	64485282	R0760:B11
1113	64485283	R0760:B12
1114	64485284	R0760:C01
1115	64485285	R0760:C02
1116	64485286	R0760:C03
1117	64485287	R0760:C04
1118	64485288	R0760:C05
1119	64485289	R0760:C06
1120	64485290	R0760:C07
1121	64485291	R0760:C08
1122	64485292	R0760:C09
1123	64485293	R0760:C10
1124	64485294	R0760:C11
1125	64485295	R0760:C12
1126	64485297	R0760:D02
1127	64485298	R0760:D03
1128	64485299	R0760:D04
1129	64485300	R0760:D05
1130	64485303	R0760:D08
1131	64485306	R0760:D11
1132	64485307	R0760:D12
1133	64485308	R0760:E01
1134	64485309	R0760:E02
1135	64485310	R0760:E03
1136	64485311	R0760:E04
1137	64485312	R0760:E05
1138	64485313	R0760:E06
1139	64485314	R0760:E07
1140	64485315	R0760:E08
1141	64485316	R0760:E09
1142	64485317	R0760:E10
1143	64485318	R0760:E11
1144	64485319	R0760:E12
1145	64485320	R0760:F01
1146	64485321	R0760:F02
1147	64485322	R0760:F03
1148	64485323	R0760:F04
1149	64485324	R0760:F05

SEQ ID NO:	Clone ID	Description
1150	64485325	R0760:F06
1151	64485326	R0760:F07
1152	64485327	R0760:F08
1153	64485328	R0760:F09
1154	64485329	R0760:F10
1155	64485330	R0760:F11
1156	64485331	R0760:F12
1157	64485332	R0760:G01
1158	64485333	R0760:G02
1159	64485334	R0760:G03
1160	64485335	R0760:G04
1161	64485336	R0760:G05
1162	64485337	R0760:G06
1163	64485338	R0760:G07
1164	64485339	R0760:G08
1165	64485340	R0760:G09
1166	64485341	R0760:G10
1167	64485342	R0760:G11
1168	64485343	R0760:G12
1169	64485344	R0760:H01
1170	64485345	R0760:H02
1171	64485346	R0760:H03
1172	64485347	R0760:H04
1173	64485348	R0760:H05
1174	64485349	R0760:H06
1175	64485350	R0760:H07
1176	64485351	R0760:H08
1177	64485352	R0760:H09
1178	64485353	R0760:H10
1179	64485354	R0760:H11
1180	64485355	R0761:A02
1181	64485356	R0761:A03
1182	64485357	R0761:A05
1183	64485358	R0761:A06
1184	64485359	R0761:A07
1185	64485360	R0761:A08
1186	64485361	R0761:A09
1187	64485362	R0761:A10
1188	64485363	R0761:A11
1189	64485364	R0761:A12
1190	64485365	R0761:B01
1191	64485366	R0761:B02

SEQ ID NO:	Clone ID	Description
1192	64485367	R0761:B03
1193	64485368	R0761:B04
1194	64485369	R0761:B05
1195	64485370	R0761:B06
1196	64485371	R0761:B07
1197	64485372	R0761:B08
1198	64485373	R0761:B09
1199	64485374	R0761:B10
1200	64485375	R0761:B11
1201	64485376	R0761:B12
1202	64485377	R0761:C01
1203	64485378	R0761:C02
1204	64485379	R0761:C03
1205	64485380	R0761:C04
1206	64485381	R0761:C05
1207	64485382	R0761:C06
1208	64485383	R0761:C07
1209	64485384	R0761:C08
1210	64485385	R0761:C09
1211	64485386	R0761:C10
1212	64485387	R0761:C11
1213	64485388	R0761:C12
1214	64485390	R0761:D02
1215	64485391	R0761:D03
1216	64485392	R0761:D04
1217	64485394	R0761:D06
1218	64485395	R0761:D07
1219	64485396	R0761:D08
1220	64485397	R0761:D09
1221	64485399	R0761:D11
1222	64485400	R0761:D12
1223	64485402	R0761:E02
1224	64485403	R0761:E03
1225	64485404	R0761:E04
1226	64485405	R0761:E05
1227	64485406	R0761:E06
1228	64485407	R0761:E07
1229	64485408	R0761:E08
1230	64485409	R0761:E09
1231	64485410	R0761:E10
1232	64485411	R0761:E11
1233	64485412	R0761:E12

SEQ ID NO:	Clone ID	Description
1234	64485413	R0761:F01
1235	64485414	R0761:F02
1236	64485415	R0761:F03
1237	64485416	R0761:F04
1238	64485417	R0761:F05
1239	64485418	R0761:F06
1240	64485419	R0761:F07
1241	64485420	R0761:F08
1242	64485421	R0761:F09
1243	64485422	R0761:F10
1244	64485423	R0761:F11
1245	64485424	R0761:F12
1246	64485426	R0761:G02
1247	64485427	R0761:G03
1248	64485428	R0761:G04
1249	64485431	R0761:G07
1250	64485432	R0761:G08
1251	64485433	R0761:G09
1252	64485434	R0761:G10
1253	64485435	R0761:G11
1254	64485436	R0761:G12
1255	64485437	R0761:H01
1256	64485438	R0761:H02
1257	64485439	R0761:H03
1258	64485440	R0761:H04
1259	64485441	R0761:H05
1260	64485442	R0761:H06
1261	64485444	R0761:H08
1262	64485445	R0761:H09
1263	64485446	R0761:H10
1264	64485447	R0761:H11
1265	64485169	R0762:A02
1266	64485170	R0762:A03
1267	64485171	R0762:A05
1268	64485172	R0762:A06
1269	64485173	R0762:A07
1270	64485174	R0762:A08
1271	64485175	R0762:A09
1272	64485176	R0762:A10
1273	64485177	R0762:A11
1274	64485178	R0762:A12
1275	64485179	R0762:B01

SEQ ID NO:	Clone ID	Description
1276	64485180	R0762:B02
1277	64485181	R0762:B03
1278	64485182	R0762:B04
1279	64485183	R0762:B05
1280	64485184	R0762:B06
1281	64485185	R0762:B07
1282	64485186	R0762:B08
1283	64485187	R0762:B09
1284	64485188	R0762:B10
1285	64485189	R0762:B11
1286	64485191	R0762:C01
1287	64485192	R0762:C02
1288	64485193	R0762:C03
1289	64485194	R0762:C04
1290	64485195	R0762:C05
1291	64485196	R0762:C06
1292	64485197	R0762:C07
1293	64485198	R0762:C08
1294	64485199	R0762:C09
1295	64485200	R0762:C10
1296	64485201	R0762:C11
1297	64485202	R0762:C12
1298	64485203	R0762:D01
1299	64485204	R0762:D02
1300	64485205	R0762:D03
1301	64485206	R0762:D04
1302	64485207	R0762:D05
1303	64485209	R0762:D07
1304	64485210	R0762:D08
1305	64485211	R0762:D09
1306	64485212	R0762:D10
1307	64485213	R0762:D11
1308	64485214	R0762:D12
1309	64485216	R0762:E02
1310	64485217	R0762:E03
1311	64485218	R0762:E04
1312	64485219	R0762:E05
1313	64485220	R0762:E06
1314	64485221	R0762:E07
1315	64485222	R0762:E08
1316	64485223	R0762:E09
1317	64485224	R0762:E10

SEQ ID NO:	Clone ID	Description
1318	64485226	R0762:E12
1319	64485227	R0762:F01
1320	64485230	R0762:F04
1321	64485231	R0762:F05
1322	64485232	R0762:F06
1323	64485233	R0762:F07
1324	64485234	R0762:F08
1325	64485235	R0762:F09
1326	64485236	R0762:F10
1327	64485237	R0762:F11
1328	64485238	R0762:F12
1329	64485239	R0762:G01
1330	64485241	R0762:G03
1331	64485242	R0762:G04
1332	64485244	R0762:G06
1333	64485245	R0762:G07
1334	64485247	R0762:G09
1335	64485248	R0762:G10
1336	64485249	R0762:G11
1337	64485250	R0762:G12
1338	64485251	R0762:H01
1339	64485253	R0762:H03
1340	64485254	R0762:H04
1341	64485255	R0762:H05
1342	64485256	R0762:H06
1343	64485257	R0762:H07
1344	64485258	R0762:H08
1345	64485260	R0762:H10
1346	64485261	R0762:H11
1347	64485076	R0763:A02
1348	64485077	R0763:A03
1349	64485078	R0763:A05
1350	64485079	R0763:A06
1351	64485080	R0763:A07
1352	64485081	R0763:A08
1353	64485082	R0763:A09
1354	64485083	R0763:A10
1355	64485084	R0763:A11
1356	64485085	R0763:A12
1357	64485086	R0763:B01
1358	64485087	R0763:B02
1359	64485089	R0763:B04

SEQ ID NO:	Clone ID	Description
1360	64485090	R0763:B05
1361	64485091	R0763:B06
1362	64485092	R0763:B07
1363	64485093	R0763:B08
1364	64485094	R0763:B09
1365	64485095	R0763:B10
1366	64485096	R0763:B11
1367	64485097	R0763:B12
1368	64485099	R0763:C02
1369	64485100	R0763:C03
1370	64485102	R0763:C05
1371	64485104	R0763:C07
1372	64485105	R0763:C08
1373	64485106	R0763:C09
1374	64485107	R0763:C10
1375	64485108	R0763:C11
1376	64485109	R0763:C12
1377	64485111	R0763:D02
1378	64485112	R0763:D03
1379	64485113	R0763:D04
1380	64485114	R0763:D05
1381	64485118	R0763:D09
1382	64485119	R0763:D10
1383	64485120	R0763:D11
1384	64485121	R0763:D12
1385	64485123	R0763:E02
1386	64485124	R0763:E03
1387	64485125	R0763:E04
1388	64485126	R0763:E05
1389	64485127	R0763:E06
1390	64485128	R0763:E07
1391	64485129	R0763:E08
1392	64485130	R0763:E09
1393	64485131	R0763:E10
1394	64485132	R0763:E11
1395	64485133	R0763:E12
1396	64485134	R0763:F01
1397	64485135	R0763:F02
1398	64485136	R0763:F03
1399	64485137	R0763:F04
1400	64485138	R0763:F05
1401	64485139	R0763:F06

SEQ ID NO:	Clone ID	Description
1402	64485140	R0763:F07
1403	64485141	R0763:F08
1404	64485142	R0763:F09
1405	64485143	R0763:F10
1406	64485144	R0763:F11
1407	64485145	R0763:F12
1408	64485147	R0763:G02
1409	64485148	R0763:G03
1410	64485149	R0763:G04
1411	64485150	R0763:G05
1412	64485151	R0763:G06
1413	64485152	R0763:G07
1414	64485153	R0763:G08
1415	64485154	R0763:G09
1416	64485155	R0763:G10
1417	64485156	R0763:G11
1418	64485157	R0763:G12
1419	64485158	R0763:H01
1420	64485159	R0763:H02
1421	64485160	R0763:H03
1422	64485161	R0763:H04
1423	64485162	R0763:H05
1424	64485163	R0763:H06
1425	64485164	R0763:H07
1426	64485165	R0763:H08
1427	64485166	R0763:H09
1428	64485167	R0763:H10
1429	64485168	R0763:H11
1430	64484983	R0764:A02
1431	64484984	R0764:A03
1432	64484985	R0764:A05
1433	64484986	R0764:A06
1434	64484987	R0764:A07
1435	64484988	R0764:A08
1436	64484990	R0764:A10
1437	64484991	R0764:A11
1438	64484992	R0764:A12
1439	64484993	R0764:B01
1440	64484994	R0764:B02
1441	64484995	R0764:B03
1442	64484996	R0764:B04
1443	64484997	R0764:B05

SEQ ID NO:	Clone ID	Description
1444	64484998	R0764:B06
1445	64484999	R0764:B07
1446	64485000	R0764:B08
1447	64485001	R0764:B09
1448	64485002	R0764:B10
1449	64485004	R0764:B12
1450	64485005	R0764:C01
1451	64485006	R0764:C02
1452	64485007	R0764:C03
1453	64485008	R0764:C04
1454	64485009	R0764:C05
1455	64485010	R0764:C06
1456	64485011	R0764:C07
1457	64485012	R0764:C08
1458	64485013	R0764:C09
1459	64485014	R0764:C10
1460	64485015	R0764:C11
1461	64485016	R0764:C12
1462	64485017	R0764:D01
1463	64485018	R0764:D02
1464	64485019	R0764:D03
1465	64485021	R0764:D05
1466	64485022	R0764:D06
1467	64485023	R0764:D07
1468	64485024	R0764:D08
1469	64485025	R0764:D09
1470	64485027	R0764:D11
1471	64485028	R0764:D12
1472	64485029	R0764:E01
1473	64485030	R0764:E02
1474	64485031	R0764:E03
1475	64485032	R0764:E04
1476	64485033	R0764:E05
1477	64485034	R0764:E06
1478	64485035	R0764:E07
1479	64485036	R0764:E08
1480	64485037	R0764:E09
1481	64485038	R0764:E10
1482	64485039	R0764:E11
1483	64485040	R0764:E12
1484	64485041	R0764:F01
1485	64485042	R0764:F02

SEQ ID NO:	Clone ID	Description
1486	64485043	R0764:F03
1487	64485044	R0764:F04
1488	64485045	R0764:F05
1489	64485046	R0764:F06
1490	64485047	R0764:F07
1491	64485048	R0764:F08
1492	64485049	R0764:F09
1493	64485050	R0764:F10
1494	64485051	R0764:F11
1495	64485052	R0764:F12
1496	64485053	R0764:G01
1497	64485054	R0764:G02
1498	64485056	R0764:G04
1499	64485057	R0764:G05
1500	64485058	R0764:G06
1501	64485059	R0764:G07
1502	64485061	R0764:G09
1503	64485062	R0764:G10
1504	64485063	R0764:G11
1505	64485065	R0764:H01
1506	64485066	R0764:H02
1507	64485067	R0764:H03
1508	64485068	R0764:H04
1509	64485069	R0764:H05
1510	64485070	R0764:H06
1511	64485071	R0764:H07
1512	64485072	R0764:H08
1513	64485073	R0764:H09
1514	64485074	R0764:H10
1515	64485075	R0764:H11
1516	64567134	R0765:A02
1517	64567135	R0765:A03
1518	64567136	R0765:A05
1519	64567137	R0765:A06
1520	64567138	R0765:A07
1521	64567139	R0765:A08
1522	64567140	R0765:A09
1523	64567141	R0765:A10
1524	64567142	R0765:A11
1525	64567143	R0765:A12
1526	64567144	R0765:B01
1527	64567145	R0765:B02

SEQ ID NO:	Clone ID	Description
1528	64567146	R0765:B03
1529	64567147	R0765:B04
1530	64567148	R0765:B05
1531	64567149	R0765:B06
1532	64567150	R0765:B07
1533	64567152	R0765:B09
1534	64567153	R0765:B10
1535	64567154	R0765:B11
1536	64567155	R0765:B12
1537	64567156	R0765:C01
1538	64567157	R0765:C02
1539	64567158	R0765:C03
1540	64567159	R0765:C04
1541	64567160	R0765:C05
1542	64567161	R0765:C06
1543	64567162	R0765:C07
1544	64567163	R0765:C08
1545	64567164	R0765:C09
1546	64567165	R0765:C10
1547	64567166	R0765:C11
1548	64567168	R0765:D01
1549	64567169	R0765:D02
1550	64567171	R0765:D04
1551	64567172	R0765:D05
1552	64567174	R0765:D07
1553	64567175	R0765:D08
1554	64567176	R0765:D09
1555	64567177	R0765:D10
1556	64567178	R0765:D11
1557	64567179	R0765:D12
1558	64567180	R0765:E01
1559	64567181	R0765:E02
1560	64567182	R0765:E03
1561	64567183	R0765:E04
1562	64567184	R0765:E05
1563	64567185	R0765:E06
1564	64567186	R0765:E07
1565	64567187	R0765:E08
1566	64567188	R0765:E09
1567	64567189	R0765:E10
1568	64567190	R0765:E11
1569	64567191	R0765:E12

SEQ ID NO:	Clone ID	Description
1570	64567192	R0765:F01
1571	64567193	R0765:F02
1572	64567194	R0765:F03
1573	64567195	R0765:F04
1574	64567196	R0765:F05
1575	64567197	R0765:F06
1576	64567198	R0765:F07
1577	64567199	R0765:F08
1578	64567200	R0765:F09
1579	64567201	R0765:F10
1580	64567203	R0765:F12
1581	64567204	R0765:G01
1582	64567205	R0765:G02
1583	64567206	R0765:G03
1584	64567207	R0765:G04
1585	64567208	R0765:G05
1586	64567210	R0765:G07
1587	64567211	R0765:G08
1588	64567212	R0765:G09
1589	64567213	R0765:G10
1590	64567214	R0765:G11
1591	64567215	R0765:G12
1592	64567216	R0765:H01
1593	64567217	R0765:H02
1594	64567218	R0765:H03
1595	64567219	R0765:H04
1596	64567220	R0765:H05
1597	64567221	R0765:H06
1598	64567222	R0765:H07
1599	64567223	R0765:H08
1600	64567224	R0765:H09
1601	64567225	R0765:H10
1602	64567226	R0765:H11
1603	64567041	R0766:A02
1604	64567042	R0766:A03
1605	64567043	R0766:A05
1606	64567044	R0766:A06
1607	64567045	R0766:A07
1608	64567046	R0766:A08
1609	64567047	R0766:A09
1610	64567048	R0766:A10
1611	64567049	R0766:A11

SEQ ID NO:	Clone ID	Description
1612	64567050	R0766:A12
1613	64567051	R0766:B01
1614	64567052	R0766:B02
1615	64567053	R0766:B03
1616	64567054	R0766:B04
1617	64567055	R0766:B05
1618	64567057	R0766:B07
1619	64567058	R0766:B08
1620	64567059	R0766:B09
1621	64567061	R0766:B11
1622	64567063	R0766:C01
1623	64567064	R0766:C02
1624	64567065	R0766:C03
1625	64567066	R0766:C04
1626	64567067	R0766:C05
1627	64567068	R0766:C06
1628	64567069	R0766:C07
1629	64567070	R0766:C08
1630	64567071	R0766:C09
1631	64567072	R0766:C10
1632	64567073	R0766:C11
1633	64567074	R0766:C12
1634	64567075	R0766:D01
1635	64567076	R0766:D02
1636	64567077	R0766:D03
1637	64567079	R0766:D05
1638	64567080	R0766:D06
1639	64567081	R0766:D07
1640	64567083	R0766:D09
1641	64567084	R0766:D10
1642	64567085	R0766:D11
1643	64567086	R0766:D12
1644	64567087	R0766:E01
1645	64567088	R0766:E02
1646	64567089	R0766:E03
1647	64567090	R0766:E04
1648	64567091	R0766:E05
1649	64567092	R0766:E06
1650	64567093	R0766:E07
1651	64567094	R0766:E08
1652	64567095	R0766:E09
1653	64567096	R0766:E10

SEQ ID NO:	Clone ID	Description
1654	64567097	R0766:E11
1655	64567098	R0766:E12
1656	64567099	R0766:F01
1657	64567100	R0766:F02
1658	64567101	R0766:F03
1659	64567102	R0766:F04
1660	64567103	R0766:F05
1661	64567105	R0766:F07
1662	64567106	R0766:F08
1663	64567107	R0766:F09
1664	64567108	R0766:F10
1665	64567109	R0766:F11
1666	64567111	R0766:G01
1667	64567112	R0766:G02
1668	64567113	R0766:G03
1669	64567114	R0766:G04
1670	64567115	R0766:G05
1671	64567116	R0766:G06
1672	64567117	R0766:G07
1673	64567118	R0766:G08
1674	64567119	R0766:G09
1675	64567120	R0766:G10
1676	64567121	R0766:G11
1677	64567122	R0766:G12
1678	64567123	R0766:H01
1679	64567124	R0766:H02
1680	64567125	R0766:H03
1681	64567126	R0766:H04
1682	64567127	R0766:H05
1683	64567128	R0766:H06
1684	64567129	R0766:H07
1685	64567130	R0766:H08
1686	64567132	R0766:H10
1687	64567133	R0766:H11
1688	64532232	R0767:A02
1689	64532233	R0767:A03
1690	64532236	R0767:A07
1691	64532238	R0767:A09
1692	64532239	R0767:A10
1693	64532240	R0767:A11
1694	64532241	R0767:A12
1695	64532242	R0767:B01

SEQ ID NO:	Clone ID	Description
1696	64532243	R0767:B02
1697	64532244	R0767:B03
1698	64532245	R0767:B04
1699	64532246	R0767:B05
1700	64532247	R0767:B06
1701	64532248	R0767:B07
1702	64532251	R0767:B10
1703	64532253	R0767:B12
1704	64532254	R0767:C01
1705	64532255	R0767:C02
1706	64532256	R0767:C03
1707	64532257	R0767:C04
1708	64532258	R0767:C05
1709	64532259	R0767:C06
1710	64532260	R0767:C07
1711	64532262	R0767:C09
1712	64532263	R0767:C10
1713	64532264	R0767:C11
1714	64532266	R0767:D01
1715	64532267	R0767:D02
1716	64532268	R0767:D03
1717	64532269	R0767:D04
1718	64532270	R0767:D05
1719	64532271	R0767:D06
1720	64532272	R0767:D07
1721	64532274	R0767:D09
1722	64532275	R0767:D10
1723	64532276	R0767:D11
1724	64532277	R0767:D12
1725	64532278	R0767:E01
1726	64532279	R0767:E02
1727	64532281	R0767:E04
1728	64532282	R0767:E05
1729	64532283	R0767:E06
1730	64532284	R0767:E07
1731	64532286	R0767:E09
1732	64532288	R0767:E11
1733	64532289	R0767:E12
1734	64532290	R0767:F01
1735	64532291	R0767:F02
1736	64532292	R0767:F03
1737	64532293	R0767:F04

SEQ ID NO:	Clone ID	Description
1738	64532296	R0767:F07
1739	64532299	R0767:F10
1740	64532300	R0767:F11
1741	64532301	R0767:F12
1742	64532302	R0767:G01
1743	64532303	R0767:G02
1744	64532304	R0767:G03
1745	64532305	R0767:G04
1746	64532306	R0767:G05
1747	64532307	R0767:G06
1748	64532308	R0767:G07
1749	64532311	R0767:G10
1750	64532312	R0767:G11
1751	64532314	R0767:H01
1752	64532315	R0767:H02
1753	64532316	R0767:H03
1754	64532317	R0767:H04
1755	64532318	R0767:H05
1756	64532319	R0767:H06
1757	64532320	R0767:H07
1758	64532322	R0767:H09
1759	64532323	R0767:H10
1760	64532324	R0767:H11
1761	64566671	R0768:A05
1762	64566672	R0768:A06
1763	64566673	R0768:A07
1764	64566674	R0768:A08
1765	64566675	R0768:A09
1766	64566676	R0768:A10
1767	64566677	R0768:A11
1768	64566679	R0768:B01
1769	64566680	R0768:B02
1770	64566682	R0768:B04
1771	64566683	R0768:B05
1772	64566684	R0768:B06
1773	64566685	R0768:B07
1774	64566686	R0768:B08
1775	64566688	R0768:B10
1776	64566689	R0768:B11
1777	64566690	R0768:B12
1778	64566691	R0768:C01
1779	64566692	R0768:C02

SEQ ID NO:	Clone ID	Description
1780	64566693	R0768:C03
1781	64566694	R0768:C04
1782	64566695	R0768:C05
1783	64566696	R0768:C06
1784	64566697	R0768:C07
1785	64566698	R0768:C08
1786	64566699	R0768:C09
1787	64566700	R0768:C10
1788	64566701	R0768:C11
1789	64566702	R0768:C12
1790	64566703	R0768:D01
1791	64566704	R0768:D02
1792	64566706	R0768:D04
1793	64566707	R0768:D05
1794	64566708	R0768:D06
1795	64566710	R0768:D08
1796	64566711	R0768:D09
1797	64566712	R0768:D10
1798	64566713	R0768:D11
1799	64566716	R0768:E02
1800	64566718	R0768:E04
1801	64566719	R0768:E05
1802	64566720	R0768:E06
1803	64566721	R0768:E07
1804	64566722	R0768:E08
1805	64566723	R0768:E09
1806	64566725	R0768:E11
1807	64566727	R0768:F01
1808	64566728	R0768:F02
1809	64566729	R0768:F03
1810	64566730	R0768:F04
1811	64566731	R0768:F05
1812	64566732	R0768:F06
1813	64566733	R0768:F07
1814	64566734	R0768:F08
1815	64566735	R0768:F09
1816	64566737	R0768:F11
1817	64566738	R0768:F12
1818	64566741	R0768:G03
1819	64566742	R0768:G04
1820	64566743	R0768:G05
1821	64566744	R0768:G06

SEQ ID NO:	Clone ID	Description
1822	64566745	R0768:G07
1823	64566746	R0768:G08
1824	64566747	R0768:G09
1825	64566748	R0768:G10
1826	64566750	R0768:G12
1827	64566751	R0768:H01
1828	64566753	R0768:H03
1829	64566754	R0768:H04
1830	64566755	R0768:H05
1831	64566756	R0768:H06
1832	64566757	R0768:H07
1833	64566758	R0768:H08
1834	64566759	R0768:H09
1835	64566760	R0768:H10
1836	64566761	R0768:H11
1837	64532418	R0769:A02
1838	64532419	R0769:A03
1839	64532421	R0769:A06
1840	64532422	R0769:A07
1841	64532423	R0769:A08
1842	64532424	R0769:A09
1843	64532426	R0769:A11
1844	64532429	R0769:B02
1845	64532431	R0769:B04
1846	64532432	R0769:B05
1847	64532433	R0769:B06
1848	64532434	R0769:B07
1849	64532435	R0769:B08
1850	64532436	R0769:B09
1851	64532437	R0769:B10
1852	64532439	R0769:B12
1853	64532440	R0769:C01
1854	64532441	R0769:C02
1855	64532442	R0769:C03
1856	64532443	R0769:C04
1857	64532444	R0769:C05
1858	64532445	R0769:C06
1859	64532446	R0769:C07
1860	64532448	R0769:C09
1861	64532449	R0769:C10
1862	64532450	R0769:C11
1863	64532451	R0769:C12

SEQ ID NO:	Clone ID	Description
1864	64532452	R0769:D01
1865	64532455	R0769:D04
1866	64532456	R0769:D05
1867	64532459	R0769:D08
1868	64532460	R0769:D09
1869	64532461	R0769:D10
1870	64532462	R0769:D11
1871	64532463	R0769:D12
1872	64532464	R0769:E01
1873	64532465	R0769:E02
1874	64532466	R0769:E03
1875	64532467	R0769:E04
1876	64532468	R0769:E05
1877	64532469	R0769:E06
1878	64532471	R0769:E08
1879	64532473	R0769:E10
1880	64532474	R0769:E11
1881	64532475	R0769:E12
1882	64532476	R0769:F01
1883	64532477	R0769:F02
1884	64532478	R0769:F03
1885	64532479	R0769:F04
1886	64532480	R0769:F05
1887	64532482	R0769:F07
1888	64532483	R0769:F08
1889	64532485	R0769:F10
1890	64532486	R0769:F11
1891	64532487	R0769:F12
1892	64532488	R0769:G01
1893	64532489	R0769:G02
1894	64532490	R0769:G03
1895	64532491	R0769:G04
1896	64532492	R0769:G05
1897	64532493	R0769:G06
1898	64532494	R0769:G07
1899	64532495	R0769:G08
1900	64532496	R0769:G09
1901	64532497	R0769:G10
1902	64532498	R0769:G11
1903	64532499	R0769:G12
1904	64532500	R0769:H01
1905	64532501	R0769:H02

SEQ ID NO:	Clone ID	Description
1906	64532502	R0769:H03
1907	64532504	R0769:H05
1908	64532505	R0769:H06
1909	64532506	R0769:H07
1910	64532507	R0769:H08
1911	64532509	R0769:H10
1912	64532510	R0769:H11
1913	64532512	R0770:A03
1914	64532513	R0770:A05
1915	64532514	R0770:A06
1916	64532515	R0770:A07
1917	64532516	R0770:A08
1918	64532517	R0770:A09
1919	64532518	R0770:A10
1920	64532519	R0770:A11
1921	64532520	R0770:A12
1922	64532521	R0770:B01
1923	64532522	R0770:B02
1924	64532523	R0770:B03
1925	64532525	R0770:B05
1926	64532526	R0770:B06
1927	64532527	R0770:B07
1928	64532528	R0770:B08
1929	64532529	R0770:B09
1930	64532530	R0770:B10
1931	64532531	R0770:B11
1932	64532532	R0770:B12
1933	64532533	R0770:C01
1934	64532534	R0770:C02
1935	64532535	R0770:C03
1936	64532536	R0770:C04
1937	64532537	R0770:C05
1938	64532538	R0770:C06
1939	64532539	R0770:C07
1940	64532540	R0770:C08
1941	64532541	R0770:C09
1942	64532542	R0770:C10
1943	64532543	R0770:C11
1944	64532544	R0770:C12
1945	64532545	R0770:D01
1946	64532546	R0770:D02
1947	64532547	R0770:D03

SEQ ID NO:	Clone ID	Description
1948	64532548	R0770:D04
1949	64532549	R0770:D05
1950	64532550	R0770:D06
1951	64532552	R0770:D08
1952	64532553	R0770:D09
1953	64532554	R0770:D10
1954	64532555	R0770:D11
1955	64532556	R0770:D12
1956	64532558	R0770:E02
1957	64532559	R0770:E03
1958	64532561	R0770:E05
1959	64532562	R0770:E06
1960	64532563	R0770:E07
1961	64532564	R0770:E08
1962	64532565	R0770:E09
1963	64532566	R0770:E10
1964	64532567	R0770:E11
1965	64532568	R0770:E12
1966	64532569	R0770:F01
1967	64532570	R0770:F02
1968	64532571	R0770:F03
1969	64532572	R0770:F04
1970	64532573	R0770:F05
1971	64532574	R0770:F06
1972	64532575	R0770:F07
1973	64532576	R0770:F08
1974	64532577	R0770:F09
1975	64532578	R0770:F10
1976	64532579	R0770:F11
1977	64532580	R0770:F12
1978	64532581	R0770:G01
1979	64532582	R0770:G02
1980	64532583	R0770:G03
1981	64532584	R0770:G04
1982	64532585	R0770:G05
1983	64532586	R0770:G06
1984	64532587	R0770:G07
1985	64532588	R0770:G08
1986	64532590	R0770:G10
1987	64532591	R0770:G11
1988	64532593	R0770:H01
1989	64532594	R0770:H02

SEQ ID NO:	Clone ID	Description
1990	64532595	R0770:H03
1991	64532596	R0770:H04
1992	64532597	R0770:H05
1993	64532599	R0770:H07
1994	64532600	R0770:H08
1995	64532601	R0770:H09
1996	64532602	R0770:H10
1997	64532603	R0770:H11
1998	64566855	R0771:A02
1999	64566856	R0771:A03
2000	64566857	R0771:A05
2001	64566858	R0771:A06
2002	64566859	R0771:A07
2003	64566860	R0771:A08
2004	64566861	R0771:A09
2005	64566862	R0771:A10
2006	64566863	R0771:A11
2007	64566865	R0771:B01
2008	64566866	R0771:B02
2009	64566867	R0771:B03
2010	64566868	R0771:B04
2011	64566869	R0771:B05
2012	64566870	R0771:B06
2013	64566871	R0771:B07
2014	64566872	R0771:B08
2015	64566873	R0771:B09
2016	64566874	R0771:B10
2017	64566875	R0771:B11
2018	64566876	R0771:B12
2019	64566877	R0771:C01
2020	64566878	R0771:C02
2021	64566879	R0771:C03
2022	64566880	R0771:C04
2023	64566881	R0771:C05
2024	64566882	R0771:C06
2025	64566883	R0771:C07
2026	64566884	R0771:C08
2027	64566885	R0771:C09
2028	64566886	R0771:C10
2029	64566887	R0771:C11
2030	64566888	R0771:C12
2031	64566889	R0771:D01

SEQ ID NO:	Clone ID	Description
2032	64566890	R0771:D02
2033	64566891	R0771:D03
2034	64566892	R0771:D04
2035	64566893	R0771:D05
2036	64566894	R0771:D06
2037	64566895	R0771:D100 07
2038	64566896	R0771:D08
2039	64566897	R0771:D09
2040	64566898	R0771:D10
2041	64566899	R0771:D11
2042	64566900	R0771:D12
2043	64566901	R0771:E01
2044	64566902	R0771:E02
2045	64566903	R0771:E03
2046	64566905	R0771:E05
2047	64566906	R0771:E06
2048	64566907	R0771:E07
2049	64566908	R0771:E08
2050	64566909	R0771:E09
2051	64566910	R0771:E10
2052	64566911	R0771:E11
2053	64566912	R0771:E12
2054	64566913	R0771:F01
2055	64566914	R0771:F02
2056	64566915	R0771:F03
2057	64566916	R0771:F04
2058	64566917	R0771:F05
2059	64566918	R0771:F06
2060	64566919	R0771:F07
2061	64566920	R0771:F08
2062	64566921	R0771:F09
2063	64566922	R0771:F10
2064	64566923	R0771:F11
2065	64566924	R0771:F12
2066	64566925	R0771:G01
2067	64566926	R0771:G02
2068	64566927	R0771:G03
2069	64566928	R0771:G04
2070	64566930	R0771:G06
2071	64566931	R0771:G07
2072	64566932	R0771:G08

SEQ ID NO:	Clone ID	Description
2073	64566933	R0771:G09
2074	64566934	R0771:G10
2075	64566935	R0771:G11
2076	64566936	R0771:G12
2077	64566937	R0771:H01
2078	64566938	R0771:H02
2079	64566939	R0771:H03
2080	64566940	R0771:H04
2081	64566941	R0771:H05
2082	64566942	R0771:H06
2083	64566943	R0771:H07
2084	64566944	R0771:H08
2085	64566945	R0771:H09
2086	64532325	R0772:A02
2087	64532326	R0772:A03
2088	64532327	R0772:A05
2089	64532328	R0772:A06
2090	64532329	R0772:A07
2091	64532330	R0772:A08
2092	64532331	R0772:A09
2093	64532332	R0772:A10
2094	64532333	R0772:A11
2095	64532334	R0772:A12
2096	64532335	R0772:B01
2097	64532336	R0772:B02
2098	64532338	R0772:B04
2099	64532339	R0772:B05
2100	64532340	R0772:B06
2101	64532341	R0772:B07
2102	64532342	R0772:B08
2103	64532343	R0772:B09
2104	64532344	R0772:B10
2105	64532345	R0772:B11
2106	64532346	R0772:B12
2107	64532347	R0772:C01
2108	64532348	R0772:C02
2109	64532349	R0772:C03
2110	64532350	R0772:C04
2111	64532351	R0772:C05
2112	64532352	R0772:C06
2113	64532353	R0772:C07
2114	64532354	R0772:C08

SEQ ID NO:	Clone ID	Description
2115	64532355	R0772:C09
2116	64532356	R0772:C10
2117	64532357	R0772:C11
2118	64532358	R0772:C12
2119	64532359	R0772:D01
2120	64532360	R0772:D02
2121	64532361	R0772:D03
2122	64532362	R0772:D04
2123	64532363	R0772:D05
2124	64532364	R0772:D06
2125	64532365	R0772:D07
2126	64532366	R0772:D08
2127	64532367	R0772:D09
2128	64532368	R0772:D10
2129	64532369	R0772:D11
2130	64532370	R0772:D12
2131	64532371	R0772:E01
2132	64532372	R0772:E02
2133	64532373	R0772:E03
2134	64532374	R0772:E04
2135	64532375	R0772:E05
2136	64532376	R0772:E06
2137	64532377	R0772:E07
2138	64532378	R0772:E08
2139	64532379	R0772:E09
2140	64532380	R0772:E10
2141	64532382	R0772:E12
2142	64532383	R0772:F01
2143	64532384	R0772:F02
2144	64532385	R0772:F03
2145	64532386	R0772:F04
2146	64532387	R0772:F05
2147	64532388	R0772:F06
2148	64532389	R0772:F07
2149	64532390	R0772:F08
2150	64532391	R0772:F09
2151	64532392	R0772:F10
2152	64532393	R0772:F11
2153	64532394	R0772:F12
2154	64532395	R0772:G01
2155	64532396	R0772:G02
2156	64532397	R0772:G03

SEQ ID NO:	Clone ID	Description
2157	64532398	R0772:G04
2158	64532399	R0772:G05
2159	64532401	R0772:G07
2160	64532402	R0772:G08
2161	64532403	R0772:G09
2162	64532404	R0772:G10
2163	64532405	R0772:G11
2164	64532406	R0772:G12
2165	64532407	R0772:H01
2166	64532408	R0772:H02
2167	64532409	R0772:H03
2168	64532410	R0772:H04
2169	64532411	R0772:H05
2170	64532413	R0772:H07
2171	64532414	R0772:H08
2172	64532415	R0772:H09
2173	64532416	R0772:H10
2174	64532417	R0772:H11
2175	64566297	R0773:A02
2176	64566298	R0773:A03
2177	64566299	R0773:A05
2178	64566300	R0773:A06
2179	64566301	R0773:A07
2180	64566302	R0773:A08
2181	64566303	R0773:A09
2182	64566304	R0773:A10
2183	64566305	R0773:A11
2184	64566306	R0773:A12
2185	64566307	R0773:B01
2186	64566308	R0773:B02
2187	64566310	R0773:B04
2188	64566311	R0773:B05
2189	64566312	R0773:B06
2190	64566313	R0773:B07
2191	64566314	R0773:B08
2192	64566315	R0773:B09
2193	64566316	R0773:B10
2194	64566317	R0773:B11
2195	64566318	R0773:B12
2196	64566319	R0773:C01
2197	64566320	R0773:C02
2198	64566321	R0773:C03

SEQ ID NO:	Clone ID	Description
2199	64566322	R0773:C04
2200	64566325	R0773:C07
2201	64566326	R0773:C08
2202	64566327	R0773:C09
2203	64566328	R0773:C10
2204	64566330	R0773:C12
2205	64566331	R0773:D01
2206	64566332	R0773:D02
2207	64566333	R0773:D03
2208	64566334	R0773:D04
2209	64566335	R0773:D05
2210	64566337	R0773:D07
2211	64566338	R0773:D08
2212	64566339	R0773:D09
2213	64566340	R0773:D10
2214	64566341	R0773:D11
2215	64566342	R0773:D12
2216	64566343	R0773:E01
2217	64566344	R0773:E02
2218	64566345	R0773:E03
2219	64566346	R0773:E04
2220	64566347	R0773:E05
2221	64566348	R0773:E06
2222	64566349	R0773:E07
2223	64566350	R0773:E08
2224	64566351	R0773:E09
2225	64566352	R0773:E10
2226	64566353	R0773:E11
2227	64566354	R0773:E12
2228	64566355	R0773:F01
2229	64566356	R0773:F02
2230	64566357	R0773:F03
2231	64566358	R0773:F04
2232	64566359	R0773:F05
2233	64566360	R0773:F06
2234	64566361	R0773:F07
2235	64566362	R0773:F08
2236	64566363	R0773:F09
2237	64566364	R0773:F10
2238	64566365	R0773:F11
2239	64566366	R0773:F12
2240	64566367	R0773:G01

SEQ ID NO:	Clone ID	Description
2241	64566368	R0773:G02
2242	64566369	R0773:G03
2243	64566371	R0773:G05
2244	64566372	R0773:G06
2245	64566373	R0773:G07
2246	64566374	R0773:G08
2247	64566375	R0773:G09
2248	64566376	R0773:G10
2249	64566377	R0773:G11
2250	64566378	R0773:G12
2251	64566379	R0773:H01
2252	64566380	R0773:H02
2253	64566381	R0773:H03
2254	64566382	R0773:H04
2255	64566383	R0773:H05
2256	64566385	R0773:H07
2257	64566386	R0773:H08
2258	64566388	R0773:H10
2259	64566389	R0773:H11
2260	64566576	R0774:A02
2261	64566577	R0774:A03
2262	64566578	R0774:A05
2263	64566580	R0774:A07
2264	64566581	R0774:A08
2265	64566582	R0774:A09
2266	64566583	R0774:A10
2267	64566584	R0774:A11
2268	64566585	R0774:A12
2269	64566587	R0774:B02
2270	64566588	R0774:B03
2271	64566589	R0774:B04
2272	64566590	R0774:B05
2273	64566591	R0774:B06
2274	64566592	R0774:B07
2275	64566593	R0774:B08
2276	64566594	R0774:B09
2277	64566595	R0774:B10
2278	64566596	R0774:B11
2279	64566597	R0774:B12
2280	64566598	R0774:C01
2281	64566599	R0774:C02
2282	64566601	R0774:C04

SEQ ID NO:	Clone ID	Description
2283	64566602	R0774:C05
2284	64566604	R0774:C07
2285	64566605	R0774:C08
2286	64566606	R0774:C09
2287	64566607	R0774:C10
2288	64566608	R0774:C11
2289	64566609	R0774:C12
2290	64566611	R0774:D02
2291	64566612	R0774:D03
2292	64566614	R0774:D05
2293	64566616	R0774:D07
2294	64566617	R0774:D08
2295	64566618	R0774:D09
2296	64566619	R0774:D10
2297	64566620	R0774:D11
2298	64566621	R0774:D12
2299	64566622	R0774:E01
2300	64566623	R0774:E02
2301	64566624	R0774:E03
2302	64566625	R0774:E04
2303	64566626	R0774:E05
2304	64566627	R0774:E06
2305	64566628	R0774:E07
2306	64566629	R0774:E08
2307	64566630	R0774:E09
2308	64566631	R0774:E10
2309	64566632	R0774:E11
2310	64566633	R0774:E12
2311	64566634	R0774:F01
2312	64566635	R0774:F02
2313	64566636	R0774:F03
2314	64566637	R0774:F04
2315	64566638	R0774:F05
2316	64566639	R0774:F06
2317	64566640	R0774:F07
2318	64566641	R0774:F08
2319	64566642	R0774:F09
2320	64566643	R0774:F10
2321	64566645	R0774:F12
2322	64566646	R0774:G01
2323	64566647	R0774:G02
2324	64566648	R0774:G03

SEQ ID NO:	Clone ID	Description
2325	64566649	R0774:G04
2326	64566651	R0774:G06
2327	64566652	R0774:G07
2328	64566653	R0774:G08
2329	64566654	R0774:G09
2330	64566655	R0774:G10
2331	64566656	R0774:G11
2332	64566657	R0774:G12
2333	64566658	R0774:H01
2334	64566659	R0774:H02
2335	64566660	R0774:H03
2336	64566661	R0774:H04
2337	64566662	R0774:H05
2338	64566663	R0774:H06
2339	64566664	R0774:H07
2340	64566665	R0774:H08
2341	64566666	R0774:H09
2342	64566667	R0774:H10
2343	64566668	R0774:H11
2344	64566762	R0775:A02
2345	64566763	R0775:A03
2346	64566764	R0775:A05
2347	64566765	R0775:A06
2348	64566766	R0775:A07
2349	64566767	R0775:A08
2350	64566768	R0775:A09
2351	64566769	R0775:A10
2352	64566770	R0775:A11
2353	64566771	R0775:A12
2354	64566772	R0775:B01
2355	64566773	R0775:B02
2356	64566774	R0775:B03
2357	64566775	R0775:B04
2358	64566776	R0775:B05
2359	64566777	R0775:B06
2360	64566778	R0775:B07
2361	64566779	R0775:B08
2362	64566780	R0775:B09
2363	64566781	R0775:B10
2364	64566782	R0775:B11
2365	64566783	R0775:B12
2366	64566784	R0775:C01

SEQ ID NO:	Clone ID	Description
2367	64566785	R0775:C02
2368	64566786	R0775:C03
2369	64566787	R0775:C04
2370	64566788	R0775:C05
2371	64566789	R0775:C06
2372	64566790	R0775:C07
2373	64566791	R0775:C08
2374	64566792	R0775:C09
2375	64566793	R0775:C10
2376	64566794	R0775:C11
2377	64566795	R0775:C12
2378	64566796	R0775:D01
2379	64566797	R0775:D02
2380	64566799	R0775:D04
2381	64566800	R0775:D05
2382	64566801	R0775:D06
2383	64566803	R0775:D08
2384	64566804	R0775:D09
2385	64566805	R0775:D10
2386	64566806	R0775:D11
2387	64566807	R0775:D12
2388	64566808	R0775:E01
2389	64566809	R0775:E02
2390	64566810	R0775:E03
2391	64566811	R0775:E04
2392	64566812	R0775:E05
2393	64566813	R0775:E06
2394	64566814	R0775:E07
2395	64566815	R0775:E08
2396	64566816	R0775:E09
2397	64566817	R0775:E10
2398	64566818	R0775:E11
2399	64566819	R0775:E12
2400	64566820	R0775:F01
2401	64566821	R0775:F02
2402	64566822	R0775:F03
2403	64566823	R0775:F04
2404	64566824	R0775:F05
2405	64566825	R0775:F06
2406	64566826	R0775:F07
2407	64566827	R0775:F08
2408	64566828	R0775:F09

SEQ ID NO:	Clone ID	Description
2409	64566829	R0775:F10
2410	64566830	R0775:F11
2411	64566831	R0775:F12
2412	64566832	R0775:G01
2413	64566833	R0775:G02
2414	64566834	R0775:G03
2415	64566835	R0775:G04
2416	64566836	R0775:G05
2417	64566837	R0775:G06
2418	64566838	R0775:G07
2419	64566839	R0775:G08
2420	64566840	R0775:G09
2421	64566841	R0775:G10
2422	64566843	R0775:G12
2423	64566844	R0775:H01
2424	64566846	R0775:H03
2425	64566847	R0775:H04
2426	64566848	R0775:H05
2427	64566849	R0775:H06
2428	64566850	R0775:H07
2429	64566851	R0775:H08
2430	64566852	R0775:H09
2431	64566853	R0775:H10
2432	64566854	R0775:H11
2433	64794122	R0776:A02
2434	64794123	R0776:A03
2435	64794124	R0776:A05
2436	64794125	R0776:A06
2437	64794126	R0776:A07
2438	64794127	R0776:A08
2439	64794128	R0776:A09
2440	64794129	R0776:A10
2441	64794130	R0776:A11
2442	64794131	R0776:A12
2443	64794132	R0776:B01
2444	64794133	R0776:B02
2445	64794134	R0776:B03
2446	64794135	R0776:B04
2447	64794136	R0776:B05
2448	64794137	R0776:B06
2449	64794138	R0776:B07
2450	64794139	R0776:B08

SEQ ID NO:	Clone ID	Description
2451	64794140	R0776:B09
2452	64794141	R0776:B10
2453	64794142	R0776:B11
2454	64794143	R0776:B12
2455	64794144	R0776:C01
2456	64794145	R0776:C02
2457	64794146	R0776:C03
2458	64794147	R0776:C04
2459	64794148	R0776:C05
2460	64794149	R0776:C06
2461	64794150	R0776:C07
2462	64794151	R0776:C08
2463	64794152	R0776:C09
2464	64794153	R0776:C10
2465	64794154	R0776:C11
2466	64794155	R0776:C12
2467	64794156	R0776:D01
2468	64794157	R0776:D02
2469	64794158	R0776:D03
2470	64794159	R0776:D04
2471	64794160	R0776:D05
2472	64794161	R0776:D06
2473	64794162	R0776:D07
2474	64794163	R0776:D08
2475	64794164	R0776:D09
2476	64794165	R0776:D10
2477	64794166	R0776:D11
2478	64794167	R0776:D12
2479	64794168	R0776:E01
2480	64794169	R0776:E02
2481	64794170	R0776:E03
2482	64794171	R0776:E04
2483	64794172	R0776:E05
2484	64794173	R0776:E06
2485	64794174	R0776:E07
2486	64794175	R0776:E08
2487	64794176	R0776:E09
2488	64794177	R0776:E10
2489	64794178	R0776:E11
2490	64794179	R0776:E12
2491	64794180	R0776:F01
2492	64794181	R0776:F02

SEQ ID NO:	Clone ID	Description
2493	64794182	R0776:F03
2494	64794183	R0776:F04
2495	64794185	R0776:F06
2496	64794186	R0776:F07
2497	64794187	R0776:F08
2498	64794188	R0776:F09
2499	64794189	R0776:F10
2500	64794190	R0776:F11
2501	64794191	R0776:F12
2502	64794192	R0776:G01
2503	64794193	R0776:G02
2504	64794194	R0776:G03
2505	64794195	R0776:G04
2506	64794196	R0776:G05
2507	64794197	R0776:G06
2508	64794198	R0776:G07
2509	64794199	R0776:G08
2510	64794200	R0776:G09
2511	64794201	R0776:G10
2512	64794202	R0776:G11
2513	64794203	R0776:G12
2514	64794204	R0776:H01
2515	64794205	R0776:H02
2516	64794206	R0776:H03
2517	64794207	R0776:H04
2518	64794209	R0776:H06
2519	64794210	R0776:H07
2520	64794211	R0776:H08
2521	64794213	R0776:H10
2522	64794214	R0776:H11

EXAMPLE 9

IDENTIFICATION OF ADDITIONAL cDNA SEQUENCES USING PCR BASED SUBTRACTION GENERATED USING PLACENTA AS THE TESTER

This Example illustrates the identification of cDNA molecules encoding placenta specific antigens that may be over expressed in a variety tumor types. This example identifies the generation of the PPS1 library using a PCR based subtraction (as described in the previous Example). The library

was generated using placenta cDNA as the tester, with the driver cDNA composed of eight normal tissues, including lung, trachea, liver, heart, bone marrow, brain, kidney, and pancreas.

Clones generated from this library were sequenced to determine the identity of the inserts, and these sequences are disclosed in SEQ ID NOs:2523-2784 (see Table 9 for details).

Table 9

SEQ ID NO:	Clone ID	Description
2523	69708	Novel
2524	69709	Novel
2525	69710	Novel
2526	69712	Novel
2527	69713	Novel
2528	69714	Novel
2529	69716	Novel
2530	69717	Novel
2531	69719	Novel
2532	69720	Novel
2533	69721	Novel
2534	69722	Novel
2535	69723	Novel
2536	69724	Novel
2527	69726	Novel
2538	69727	Novel
2539	69728	Novel
2540	69729	Novel
2541	69731	Novel
2542	69732	Novel
2543	69733	Novel
2544	69734	Novel
2545	69735	Novel
2546	69736	Novel
2547	69737	Novel
2548	69739	Novel
2549	69741	Novel
2550	69742	Novel
2551	69743	Novel

SEQ ID NO:	Clone ID	Description
2552	69745	Novel
2553	69746	Novel
2554	69748	Novel
2555	69749	Novel
2556	69750	Novel
2557	69751	Novel
2558	69752	Novel
2559	69753	Novel
2560	69755	Novel
2561	69756	Novel
2562	69757	Novel
2563	69758	Novel
2564	69760	Novel
2565	69761	Novel
2566	69762	Novel
2567	69763	Novel
2568	69765	Novel
2569	69766	Novel
2570	69767	Novel
2571	69768	Novel
2572	69769	Novel
2573	69770	Novel
2574	69771	Novel
2575	69772	Novel
2576	69773	Novel
2577	69774	Novel
2578	69775	Novel
2579	69776	Novel
2580	69777	Novel
2581	69778	Novel
2582	69779	Novel
2583	69780	Novel
2584	69781	Novel
2585	69782	Novel
2586	69783	Novel
2587	69784	Novel
2588	69785	Novel
2589	69786	Novel
2590	69788	Novel
2591	69789	Novel
2592	69790	Novel
2593	69791	Novel

SEQ ID NO:	Clone ID	Description
2594	69792	Novel
2595	69793	Novel
2596	69794	Novel
2597	69795	Novel
2598	69796	Novel
2699	69797	Novel
2600	69798	Novel
2601	69799	Novel
2602	69800	Novel
2603	69801	Novel
2604	69802	Novel
2605	69803	Novel
2606	64566390	R0804:A02
2607	64566391	R0804:A03
2608	64566392	R0804:A05
2609	64566393	R0804:A06
2610	64566394	R0804:A07
2611	64566395	R0804:A08
2612	64566396	R0804:A09
2613	64566398	R0804:A11
2614	64566399	R0804:A12
2615	64566400	R0804:B01
2616	64566401	R0804:B02
2617	64566403	R0804:B04
1618	64566404	R0804:B05
2619	64566406	R0804:B07
2620	64566407	R0804:B08
2621	64566409	R0804:B10
2622	64566410	R0804:B11
2623	64566413	R0804:C02
2624	64566414	R0804:C03
2625	64566415	R0804:C04
2626	64566419	R0804:C08
2627	64566420	R0804:C09
2628	64566421	R0804:C10
2629	64566426	R0804:D03
2630	64566428	R0804:D05
2631	64566430	R0804:D07
2632	64566431	R0804:D08
2633	64566432	R0804:D09
2634	64566433	R0804:D10
2635	64566434	R0804:D11

SEQ ID NO:	Clone ID	Description
2636	64566436	R0804:E01
2637	64566437	R0804:E02
2638	64566438	R0804:E03
2639	64566439	R0804:E04
2640	64566442	R0804:E07
2641	64566444	R0804:E09
2642	64566445	R0804:E10
2643	64566446	R0804:E11
2644	64566447	R0804:E12
2645	64566451	R0804:F04
2646	64566452	R0804:F05
2647	64566454	R0804:F07
2648	64566457	R0804:F10
2649	64566458	R0804:F11
2650	64566460	R0804:G01
2651	64566461	R0804:G02
2652	64566462	R0804:G03
2653	64566464	R0804:G05
2654	64566465	R0804:G06
2655	64566466	R0804:G07
2656	64566468	R0804:G09
2657	64566469	R0804:G10
2658	64566470	R0804:G11
2659	64566471	R0804:G12
2660	64566473	R0804:H02
2661	64566474	R0804:H03
2662	64566475	R0804:H04
2663	64566476	R0804:H05
2664	64566477	R0804:H06
2665	64566480	R0804:H09
2666	64566482	R0804:H11
2667	64566949	R0805:A03
2668	64566950	R0805:A05
2669	64566951	R0805:A06
2670	64566952	R0805:A07
2671	64566953	R0805:A08
2672	64566954	R0805:A09
2673	64566955	R0805:A10
2674	64566957	R0805:A12
2675	64566958	R0805:B01
2676	64566959	R0805:B02
2677	64566960	R0805:B03

SEQ ID NO:	Clone ID	Description
2678	64566961	R0805:B04
2679	64566964	R0805:B07
2680	64566965	R0805:B08
2681	64566966	R0805:B09
2682	64566967	R0805:B10
2683	64566968	R0805:B11
2684	64566969	R0805:B12
2685	64566970	R0805:C01
2686	64566971	R0805:C02
2687	64566974	R0805:C05
2688	64566980	R0805:C11
2689	64566988	R0805:D07
2690	64566989	R0805:D08
2691	64566993	R0805:D12
2692	64566994	R0805:E01
2693	64566996	R0805:E03
2694	64566999	R0805:E06
2695	64567000	R0805:E07
2696	64567002	R0805:E09
2697	64567006	R0805:F01
2698	64567007	R0805:F02
2699	64567008	R0805:F03
2700	64567013	R0805:F08
2701	64567014	R0805:F09
2702	64567016	R0805:F11
2703	64567017	R0805:F12
2704	64567018	R0805:G01
2705	64567023	R0805:G06
2706	64567024	R0805:G07
2707	64567025	R0805:G08
2708	64567026	R0805:G09
2709	64567027	R0805:G10
2710	64567028	R0805:G11
2711	64567029	R0805:G12
2712	64567030	R0805:H01
2713	64567032	R0805:H03
2714	64567033	R0805:H04
2715	64567034	R0805:H05
2716	64567035	R0805:H06
2717	64567036	R0805:H07
2718	64567037	R0805:H08
2719	64567038	R0805:H09

SEQ ID NO:	Clone ID	Description
2720	64567039	R0805:H10
2721	64566484	R0806:A03
2722	64566485	R0806:A05
2723	64566487	R0806:A07
2724	64566489	R0806:A09
2725	64566490	R0806:A10
2726	64566491	R0806:A11
2727	64566497	R0806:B05
2728	64566499	R0806:B07
2729	64566500	R0806:B08
2730	64566502	R0806:B10
2731	64566503	R0806:B11
2732	64566504	R0806:B12
2733	64566505	R0806:C01
2734	64566506	R0806:C02
2735	64566507	R0806:C03
2736	64566508	R0806:C04
2737	64566510	R0806:C06
2738	64566511	R0806:C07
2739	64566513	R0806:C09
2740	64566514	R0806:C10
2741	64566515	R0806:C11
2742	64566516	R0806:C12
2743	64566518	R0806:D02
2744	64566519	R0806:D03
2745	64566520	R0806:D04
2746	64566522	R0806:D06
2747	64566524	R0806:D08
2748	64566525	R0806:D09
2749	64566526	R0806:D10
2750	64566527	R0806:D11
2751	64566528	R0806:D12
2752	64566529	R0806:E01
2753	64566530	R0806:E02
2754	64566531	R0806:E03
2755	64566532	R0806:E04
2756	64566534	R0806:E06
2757	64566535	R0806:E07
2758	64566536	R0806:E08
2759	64566537	R0806:E09
2760	64566538	R0806:E10
2761	64566539	R0806:E11

SEQ ID NO:	Clone ID	Description
2762	64566540	R0806:E12
2763	64566541	R0806:F01
2764	64566542	R0806:F02
2765	64566546	R0806:F06
2766	64566547	R0806:F07
2767	64566549	R0806:F09
2768	64566550	R0806:F10
2769	64566551	R0806:F11
2770	64566554	R0806:G02
2771	64566555	R0806:G03
2772	64566556	R0806:G04
2773	64566558	R0806:G06
2774	64566563	R0806:G11
2775	64566568	R0806:H04
2776	64566570	R0806:H06
2777	64566571	R0806:H07
2778	64566574	R0806:H10
2779	64566575	R0806:H11
2780	64782431	R0807:A08
2781	64782437	R0807:B02
2782	64782446	R0807:B11
2783	64782447	R0807:B12
2784	64782448	R0807:C01

EXAMPLE 10

CHARACTERIZATION OF THE ONCOFETAL CLONE MMP11

Clones 55218 (SEQ ID NO:23) and 56394 (SEQ ID NO:57) were identified from the S2 subtraction library, as described in Example 1. These two clones were determined to contain partial cDNA sequences that showed sequence identity to the human stromelysin-3 gene (MMP11), a member of the matrix metalloproteinase (MMP) family of proteins. Here the full-length cDNA and amino acid sequences of MMP11 are disclosed in SEQ ID NOs:2785 and 2786, respectively.

MMPs are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling. This family of proteins has also been implicated in

disease processes, such as arthritis and cancer metastasis. The majority of MMPs are secreted as inactive proproteins, which are activated when cleaved by extracellular proteinases. There are two types of MMPs, membrane bound proteins including MMP14, MMP17, and MMP24, and secreted MMPs which include MMP1, MMP10, MMP13, MMP12, MMP3, MMP2, 20, MMP21, MMP7, and MMP26.

Using microarray analysis the expression profile of MMP11 or Stromelysin-3 was evaluated in both normal and tumor tissues. Using this approach, cDNA sequences were PCR amplified and their mRNA expression profiles in tumor and normal tissues were examined using cDNA microarray technology essentially as described (Shena, M. et al., *Science 270*:467-70 (1995)). In brief, sequences of interest were arrayed onto glass slides as multiple replicas. Each chip was then hybridized with a pair of cDNA probes that were fluorescenctly labeled with Cy3 and Cy5, respectively. Typically, 1μg of polyA$^+$ RNA was used to generate each cDNA probe. After hybridization, the chips were scanned and the fluorescence intensity recorded in both Cy3 and Cy5 channels.

Compared to normal tissues, microarray analysis demonstrated that MMP11 had an average 4-fold over-expression in both bladder tumors (n=5) and pancreatic tumors (n=3), a 2.5-fold over-expression in hematological tumors (n=4), as well as elevated expression in liver and stomach cancers.

To further evaluate the tissue expression of MPP11, Real-time PCR analysis was performed. Real-time PCR (*see* Gibson et al., *Genome Research 6*:995-1001, 1996; Heid et al., *Genome Research 6*:986-994, 1996) is a technique that evaluates the level of PCR product accumulation during amplification. This technique permits quantitative evaluation of mRNA levels in multiple samples. Briefly, mRNA was extracted from tumor and normal tissue and cDNA was prepared using standard techniques. Real-time PCR was performed, for example, using a Perkin Elmer/Applied Biosystems (Foster City, CA) 7700 Prism instrument. Matching primers and fluorescent probes specific for MMP11 were designed and their optimal concentrations determined. To

quantitate the amount of MMP11 RNA in a sample, a standard curve was generated using a plasmid containing the MMP11 gene. Standard curves were generated using the Ct values determined in the real-time PCR, which were related to the initial cDNA concentration used in the assay. Standard dilutions ranging from 10-10^6 copies of the gene of interest are generally sufficient. In addition, a standard curve is generated for the control sequence. This permits standardization of initial RNA content of a tissue sample to the amount of control for comparison purposes.

Real-time PCR analysis correlated well with the microarray profiles. MMP11 was found to be highly over-expressed in the majority of cancers examined, including liver, stomach, pancreas, bladder, head and neck, and lung. In contrast, the expression of MMP11 was low or undetectable in all normal tissues, with the exception of placenta, which demonstrated high over-expression and fetal tissue, which showed some over-expression.

Therefore, MMP11 has been shown to have an expression pattern which makes it an excellent pan-tumor marker. In addition, it is a good vaccine candidate for multiple types of cancer including, but not limited to, liver, stomach, pancreas, bladder, and head and neck.

Other members of the MMP family of proteins were also evaluated to determine their expression profiles in both normal and tumor samples.

The full length cDNA and amino acid sequences for MMP10 or stromelysin 2 are disclosed in SEQ ID NOs:2904 and 2913, respectively. Real time analysis demonstrated that this sequence was over-expressed in kidney, and bladder tumors, as well as in head and neck tumors. Of the normal tissues tested, there was low-level expression was only seen in liver and trachea.

The full length cDNA and amino acid sequences for MMP1 or interstitial collagenase are disclosed in SEQ ID NOs:2905 and 2912, respectively. Real time analysis demonstrated that this sequence was over-expressed in kidney, stomach, and in head and neck tumors. No normal tissues showed any detectable expression of MMP1.

The full length cDNA and amino acid sequences for MMP13 or collagenase 3 are disclosed in SEQ ID NOs:2906 and 2914, respectively. Real time analysis demonstrated that this sequence was over-expressed in stomach, pancreas, bladder tumors, as well as in head and neck tumors. No normal tissues showed any detectable levels of MMP13.

The full length cDNA and amino acid sequences for MMP26 are disclosed in SEQ ID NOs:2909 and 2917, respectively. Real time analysis demonstrated that this sequence was highly over-expressed in kidney tumor. No normal tissues showed any detectable levels of MMP26.

The full length cDNA and amino acid sequences for MMP7 are disclosed in SEQ ID NOs:2908 and 2916, respectively. Real time analysis demonstrated that this sequence was over-expressed in liver, kidney, stomach, pancreas, bladder, and head and neck tumors. Low-level expression was seen only in normal pancreas and kidney.

The full length cDNA and amino acid sequences for MMP12 macrophage elastase are disclosed in SEQ ID NOs:2910 and 2918, respectively. Real time analysis demonstrated that this sequence was over-expressed in kidney, stomach, pancreas, bladder, head and neck tumors, as well as in small cell primary tumors and pooled squamous tumors.

The full length cDNA and amino acid sequences for MMP3 or stromelysin 1 are disclosed in SEQ ID NOs:2911 and 2919, respectively. Real time analysis demonstrated that this sequence was over-expressed in kidney, stomach, bladder, and head and neck tumors. Low levels of expression were found in normal bladder and esophagus.

The full length cDNA and amino acid sequences for MMP24 are disclosed in SEQ ID NOs:2907 and 2915, respectively. Real time analysis demonstrated that this sequence was over-expressed in liver, kidney, pancreas, and head and neck tumors, as well as in small cell primary tumors and pooled squamous tumors. Low levels of expression were also found in normal pancreas, brain, pituitary gland, testis, placenta, and pooled fetal tissue.

EXAMPLE 11

IDENTIFICATION OF ADDITIONAL ONCOFETAL cDNA SEQUENCES USING ELECTRONIC SUBTRACTION AND ELECTRONIC NORTHERN

Sequences disclosed here were generated using electronic subtraction. The testers included 436,379 genebins derived from 81 late stage fetal tissue libraries. The drivers included 723,535 genebins derived from 190 normal tissue libraries. The subtraction criterion was set such that only genebins that were unique to the tester libraries were retrieved. The subtraction resulted in the identification of greater than 67,000 sequences that were unique to the tester libraries. These sequences were then submitted for electronic northern blot analysis (eNorthern). The eNorthern was performed by determining the number of hits in two groups of libraries, one of which included 368 tumor tissue libraries, the other of which included 370 libraries from tumor associated or matched and diseased normal tissues. After eNorthern analysis only 6779 genebins remained, which had at least one hit in the tumor tissue libraries. Out of these genebins, 103 had 4 or more hits in the tumor tissue libraries. In addition, these sequences were shown to have a tumor:normal ratio of ≥5. cDNA sequences are disclosed in SEQ ID NOs:2787-2889 (see Table 10 for details).

Table 10

Oncofetal cDNA Sequences Over-Expressed in Tumor Tissues

SEQ ID NO:	CLONE ID	DESCRIPTION
2787	983688.1	983688.1 GI|7020592|DBJ|AK000477.1|AK000477 HOMO SAPIENS CDNA FLJ20470 FIS, CLONE KAT06815
2788	985795.1	NOVEL
2789	986838.1	NOVEL
2790	040982.7	040982.7 GI|4581562|GB|U86074.1|HSU86074 HOMO SAPIENS TESMIN MRNA, COMPLETE CDS
2791	338602.1	NOVEL

SEQ ID NO:	CLONE ID	DESCRIPTION
2792	401004.2	NOVEL
2793	442061.2	NOVEL
2794	406760.1	NOVEL
2795	454799.1	454799.1 GI|14424736|GB|BC009383.1|BC009383 HOMO SAPIENS, HYPOTHETICAL PROTEIN MGC10791, CLONE MGC:16709 IMAGE:4127918, MRNA, COMPLETE CDS
2796	032353.2	NOVEL
2797	244415.1	NOVEL
2798	332308.1	NOVEL
2799	342491.1	NOVEL
2800	407323.1	NOVEL
2801	449142.1	NOVEL
2802	011553.3	NOVEL
2803	220753.1	220753.1 GI|10432731|DBJ|AK021531.1|AK021531 HOMO SAPIENS CDNA FLJ11469 FIS, CLONE HEMBA1001658
2804	355885.1	NOVEL
2805	199702.1	199702.1 GI|11419693|REF|XM_010097.1|HOMO SAPIENS POTASSIUM VOLTAGE-GATED CHANNEL, ISK-RELATED FAMILY, MEMBER 1-LIKE (KCNE1L), MRNA
2806	005785.1	NOVEL
2807	203464.1	NOVEL
2808	404352.1	NOVEL
2809	239996.3	NOVEL
2810	259362.8	259362.8 GI|15298521|REF|XM_043136.2|HOME SAPIENS KIAA1617 PROTEIN (KIAA1617), MRNA
2811	980489.1	NOVEL
2812	408716.1	NOVEL
2813	231114.1	NOVEL
2814	030443.1	NOVEL
2815	025512.1	NOVEL

SEQ ID NO:	CLONE ID	DESCRIPTION
2816	410309.4	410309.4 GI|15305066|REF|XM_017175.3|HOMO SAPIENS DESMOCOLLIN 1 (DSC1), MRNA
2817	186692.1	NOVEL
2818	060413.1	NOVEL
2819	133555.1	133555.1 GI|13376112|REF|NM_024768.1|HOMO SAPIENS HYPOTHETICAL PROTEIN FLJ12057 (FLJ12057), MRNA
2820	209851.1	NOVEL
2821	338674.1	NOVEL
2822	346544.5	NOVEL
2823	400432.1	NOVEL
2824	075494.1	NOVEL
2825	981240.1	NOVEL
2826	239218.1	239218.1 GI|10434188|DBJ|AK022664.1|AK022664 HOMO SAPIENS CDNA FLJ12602 FIS, CLONE NT2RM4001437
2827	345913.3	NOVEL
2828	097649.1	NOVEL
2829	404257.3	404257.3 GI|12751476|REF|NM_023067.1|HOMO SAPIENS FORKHEAD TRANSCRIPTION FACTOR FOXL2 (BPES), MRNA
2830	216038.1	216038.1 GI|3483601|GB|AF086256.1\HUMZD41C11 HOMO SAPIENS FULL LENGTH INSERT CDNA CLONE ZD41C11
2831	410634.1	410634.1 GI|189659|GB|M55514.1|HUMPCC HUMAN POTASSIUM CHANNEL (HPCN2) MRNA, COMPELTE CDS
2832	440457.1	NOVEL
2833	058031.1	NOVEL
2834	207665.1	NOVEL
2835	226522.1	NOVEL
2836	202406.1	NOVEL
2837	215060.1	NOVEL
2838	133059.1	NOVEL

SEQ ID NO:	CLONE ID	DESCRIPTION
2839	233254.1	NOVEL
2840	095513.1	NOVEL
2841	004685.4	NOVEL
2842	058807.1	NOVEL
2843	402762.2	NOVEL
2844	172575.1	NOVEL
2845	003619.1	NOVEL
2846	017419.1	NOVEL
2847	171473.4	NOVEL
2848	213590.1	NOVEL
2849	237182.2	NOVEL
2850	239929.1	NOVEL
2851	337627.1	NOVEL
2852	366959.1	NOVEL
2853	391973.1	391973.1 GI|10434776|DBJ\AK023045.1|AK023045 HOMO SAPIENS CDNA FLJ12983 FIS, CLONE NT2RP3000002
2854	400216.1	NOVEL
2855	406357.1	406357.1 GI|13507323|GB|AF339785.1|AF339785 HOMO SAPIENS CLONE IMAGE:1963178, MRNA SEQUENCE
2856	407800.1	NOVEL
2857	415369.1	NOVEL
2858	016710.1	NOVEL
2859	016716.1	NOVEL
2860	342726.1	NOVEL
2861	006255.1	NOVEL
2862	297540.1	NOVEL
2863	110375.1	NOVEL
2864	329762.1	NOVEL
2865	026289.1	NOVEL
2866	044421.1	NOVEL
2867	402085.1	NOVEL

SEQ ID NO:	CLONE ID	DESCRIPTION
2868	334715.1	NOVEL
2869	360246.1	NOVEL
2870	438682.1	NOVEL
2871	105849.1	NOVEL
2872	000356.1	NOVEL
2873	013807.1	NOVEL
2874	050052.1	NOVEL
2875	207794.1	207794.1 GI|15559391|GB|BC014063.1|BC014063 HOMO SAPIENS, CLONE IMAGE:3844647, MRNA
2876	212624.1	NOVEL
2877	298825.1	NOVEL
2878	332509.1	NOVEL
2879	399838.1	NOVEL
2880	205835.1	NOVEL
2881	400510.1	NOVEL
2882	414051.1	NOVEL
2883	460590.1	NOVEL
2884	171616.1	NOVEL
2885	177824.1	NOVEL
2886	260794.1	NOVEL
2887	338503.1	NOVEL
2888	331530.1	NOVEL
2889	197089.1	NOVEL

EXAMPLE 12

MICROARRAY ANALYSIS OF CLONES IDENTIFIED FROM THE TPS1 AND PPS1 SUBTRACTION LIBRARIES.

This example describes microarray expression analysis of sequences identified using the TPS1 library (described in Example 8) and the PPS1 library (described in Example 9). A total of 2041 clones (1728 from the TPS1 library and 313 from the PPS1 library) were arrayed on Placenta/Testis

Chip 1. cDNA inserts from the clones were amplified by PCR using vector specific primers. The resulting PCR products were then sequenced in one direction. The array was then probed with probe pairs representing a variety of tumor and normal tissues, including brain, lung, heart, liver, kidney, pancreas, bone marrow, and trachea. Analysis consisted of determining the ratio of the mean hybridization signal for a particular cDNA sequence using two sets of probe groups. The determined ratio is a reflection of the over- or under-expression of a sequence within a probe population (*i.e.*, a specific tissue). The probe groups were set up to identify sequences with high differential expression in the tumor probe group compared to the normal probe group. A threshold for the ratio of over-expression between the tumor probe group and the normal tissue probe group was set at 3.0. Forty-six sequences were identified which showed a tumor group:normal group ratio of 3 or greater. These cDNA sequences, including their ratios are described in Table 11.

Table 11

Microarray Analysis Of Clones Identified From the TPS1 and PPS1 Subtraction Libraries

SEQ ID NO:	Clone ID	Ratio	Tumor	Normal	Description
1112	64485282	8.05	1.237	0.154	GAGE
1108	64485278	3.58	0.725	0.202	host cell factor 2 (AF117210)
1134	64485309	3.23	0.546	0.169	proteasome (prosome, macropain) subunit, alpha type 6 (BC002979)
1265	64485169	3.3	0.599	0.182	cDNA DKFZp434A1315
1141	64485150	5.94	1.571	0.265	PDZ-binding kinase (AF189722)
1427	64485166	3.15	0.567	0.18	RP42 protein (AF292100)
1435	64484986	5.09	0.425	0.084	kazal-type serine proteinase (M91438)

SEQ ID NO:	Clone ID	Ratio	Tumor	Normal	Description
1484	64485041	10.34	1.648	0.159	chromosome 13 (AL160033)
1528	64567146	3.43	0.789	0.23	hiwi (AF104260)
1563	64567185	3.54	0.425	0.12	desmoglein 1 (XM_008810)
1653	64567096	6.39	1.485	0.232	putative nucleoside diphosphate kinase (Y14992)
1666	64567111	14.85	2.265	0.152	testicular protein (TSPY) (M98525)
1768	64566679	6.47	0.822	0.127	testis-specific ankyrin motif containing protein (NM_019644)
1775	64566688	5.7	1.048	0.184	activator of S phase kinase (NM_006716)
1802	64566720	3.45	0.525	0.152	clone RP3-382I10 on chromosome 6q14.3-15
1825	64566748	3.82	0.583	0.152	EF-hand protein, 1 (XM_008755)
1875	64532467	3.04	1.008	0.331	p80-coilin (U06632)
1882	64532476	4.56	1.434	0.314	plasma serine protease (protein C) inhibitor (J02639)
1964	64532567	167.16	18.508	0.111	GAGE-2 protein (U19143)
1965	64532568	16.12	3.278	0.203	ropporin (AF303889)
1958	64532561	5.17	1.383	0.268	Pr22 protein
1959	64532562	101.27	27.515	0.272	GAGE
1961	64532564	4.6	0.709	0.154	transketolase-like protein (X91817)
1977	64532580	23.88	3.417	0.143	GAGE-6
1971	64532574	32.57	4.1	0.126	mitochondrial capsule selenoprotein (X89960)
1982	64532585	3.45	0.618	0.179	ubiquitin carrier protein (BC007554)
1983	64532586	3.54	0.646	0.182	leucyl/cystinyl aminopeptidase

SEQ ID NO:	Clone ID	Ratio	Tumor	Normal	Description
					(NM_005575)
2053	64566912	3.33	0.584	0.176	testis-specific ankyrin motif containing protein
2060	64566919	7.82	4.968	0.635	plasma serine protease (protein C) inhibitor (J02639)
2085	64566945	10.13	2.115	0.209	putative cell surface receptor NYD-SP8 (AY014285)
2171	64532414	3.94	1.126	0.286	Pr22 protein
2181	64566303	93.24	15.172	0.163	GAGE4
2208	64566344	71.74	33.871	0.472	GAGE-2
2263	64566580	4.26	2.342	0.55	cysteine-rich secretory protein-2/type I (X95239)
2295	64566618	5.21	3.938	0.755	cysteine-rich secretory protein-2/type I (X95239)
2299	64566622	5.47	0.837	0.153	cysteine-rich secretory protein-2/type I (X95239)
2382	64566801	5.4	1.548	0.287	KIAA1830 protein
2397	64566817	4.32	0.677	0.157	testis-specific ankyrin motif containing protein (NM_019644)
2497	64794187	4.77	0.757	0.159	Pr22 protein
2509	64794199	4.12	0.874	0.212	mitotic centromere-associated kinesin (U63743)
2518	64794209	3.68	0.515	0.14	testis-specific ankyrin motif containing protein
2652	64566462	4.02	0.912	0.227	disintegrin and metalloproteinase domain 12 (XM_005838)
2689	64566988	4.9	1.34	0.273	clone 23887 mRNA (AF052114)
2713	64567032	4.72	1.353	0.287	pregnancy specific beta-1-glycoprotein 4 (NM_002780)
2750	64566527	3.31	1.33	0.402	pregnancy-specific beta-1-glycoprotein

SEQ ID NO:	Clone ID	Ratio	Tumor	Normal	Description
2752	64566529	6.54	2.035	0.311	mitogen-responsive phosphoprotein DOC-2 (U53446)

EXAMPLE 13

REAL TIME PCR ANALYSIS OF CLONES IDENTIFIED FROM THE TPS1 AND PPS1 SUBTRACTION LIBRARIES.

Real-time PCR analysis was performed on sequences shown to be over-expressed in tumor samples when compared to normal tissues. Real-time analysis was performed essentially as described in Example 10.

Clone 64532567, initially disclosed in SEQ ID NO:1964, was previously shown, using microarray analysis, to have a tumor:normal ratio of 167.16 (see Example 12 for details). Real-time PCR analysis demonstrated that the sequence specific for clone 64532567 was over-expressed in small cell lung carcinoma and lung adenocarcinoma, with no detectable expression in any normal tissues tested including: lung, brain, pituitary gland, adrenal gland, thyroid gland, pancreas, heart, liver, skeletal muscle, stomach, kidney, small intestine, colon, bladder, esophagus, skin, salivary gland, trachea, bone marrow, resting PBMCs, spleen, lymph node, thymus and spinal cord. When SEQ ID NO;1964 was searched against publicly available databases, full length cDNA and protein sequences encoding the GAGE-2 protein were found. These sequences are disclosed in SEQ ID NOs:2890 and 2895, respectively.

Clone 64567111, also described as OF1704P, was initially disclosed in SEQ ID NO:1666, was shown using microarray analysis to have a tumor:normal ratio of 14.85 (see Example 12 for details). Real-time PCR analysis demonstrated that the sequence specific for clone 64567111 was over-expressed in lung squamous cell carcinoma, liver cancer and normal testis, with no detectable expression in any other normal tissues tested including: lung, brain, pituitary gland, adrenal gland, thyroid gland, pancreas, heart, liver, skeletal muscle, stomach, kidney, small intestine, colon, bladder, esophagus,

skin, salivary gland, trachea, bone marrow, resting PBMCs, spleen, lymph node, thymus and spinal cord. Using SEQ ID NO:1666 to search publicly available databases revealed full length cDNA and protein sequences encoding the Testis Specific Protein Y (TSPY). These sequences are disclosed in SEQ ID NOs:2891 and 2896, respectively.

Clone 64566945, also known as OF1703P, was initially disclosed in SEQ ID NO:2085, and shown using microarray analysis to have a tumor:normal ratio of 10.13 (see Example 12 for details). Real-time PCR analysis demonstrated that the sequence specific for clone 64566945 was over-expressed in small cell lung carcinoma, with low levels of expression in skeletal muscle and soft palate. There was no detectable expression in any other normal tissues tested including: lung, brain, pituitary gland, adrenal gland, thyroid gland, pancreas, heart, liver, stomach, kidney, small intestine, colon, bladder, esophagus, skin, salivary gland, trachea, bone marrow, resting PBMCs, spleen, lymph node, thymus and spinal cord. Using SEQ ID NO:2085 to search publicly available databases revealed full length cDNA and protein sequences encoding the a putative cell surface receptor, NYD-SP8. These sequences are disclosed in SEQ ID NOs:2892 and 2897, respectively.

Clone 64566529, also described as OF1705P, was initially disclosed in SEQ ID NO:2752, and shown using microarray analysis to have a tumor:normal ratio of 6.54 (see Example 12 for details). Real-time PCR analysis demonstrated that the sequence specific for clone 64566529 was over-expressed in kidney tumors and placenta, with low levels in lymph nodes. There was no detectable expression in any other normal tissues tested, including lung, brain, pituitary gland, adrenal gland, thyroid gland, pancreas, heart, liver, skeletal muscle, stomach, kidney, small intestine, colon, bladder, esophagus, skin, salivary gland, trachea, bone marrow, resting PBMCs, spleen, thymus and spinal cord. Using SEQ ID NO:2752 to search publicly available databases revealed full length cDNA and protein sequences encoding the mitogen-responsive phosphoprotein DOC-2 protein. These sequences are disclosed in SEQ ID NOs:2893 and 2898, respectively.

Clone 64567185 was initially disclosed in SEQ ID NO:1563, and shown using microarray analysis to have a tumor:normal ratio of 3.54 (see Example 12 for details). Real-time PCR analysis demonstrated that the sequence specific for clone 64567185 was over-expressed in bladder cancer, stomach, esophagus, and testis. There was no detectable expression in any other normal tissues tested including: lung, brain, pituitary gland, adrenal gland, thyroid gland, pancreas, heart, liver, skeletal muscle, kidney, small intestine, colon, bladder, skin, salivary gland, trachea, bone marrow, resting PBMCs, spleen, thymus and spinal cord. Using SEQ ID NO:1563 to search publicly available databases, revealed full length cDNA and protein sequences encoding the Desmoglein 1 protein. These sequences are disclosed in SEQ ID NOs:2894 and 2899, respectively.

EXAMPLE 14

CHARACTERIZATION OF CLONE NUMBER 410309

Clone 41039 (SEQ ID NO:2816) was identified using the eSubtraction library derived from late fetal tissues which was described in detail in Example 11. This sequence was found to encode the Desmocollin-1 protein. The Desmocollin-1 gene was found to encode two splice forms, the cDNA and protein sequences of which are disclosed in SEQ ID NOs:2900-2901 and 2902-2903, respectively.

Real-time PCR analysis of clone 410309 revealed that it was over-expressed in bladder tumors. Lower levels of expression were also detected in lung squamous cell carcinomas, head and neck carcinoma, as well as small cell carcinoma. Of the normal tissues tested, expression was only detected in normal skin.

EXAMPLE 15

FURTHER CHARACTERIZATION OF THE ONCOFETAL CLONE MMP-11

Sequences encoding the oncofetal clone MMP-11, also known as OF643S, were identified using the S2 subtraction library, essentially as

described in Example 1. The full length cDNA and amino acid sequences of which are disclosed in SEQ ID NOs:2785 and 2786, respectively. This clone was analyzed using real time PCR in Example 10, where it was identified as a pan-tumor marker. Here, the sequence was analyzed on an extended panel of pancreas tumor samples. Real-time PCR analysis revealed that MMP-11 was over-expressed in 21/22 pancreas tumor samples tested, including one pancreatic metastatic sample, with very low-level expression in normal pancreas. Low-level expression was also detected in normal testis and uterus.

EXAMPLE 16

PEPTIDE PRIMING OF T-HELPER LINES

Generation of $CD4^+$ T helper lines and identification of peptide epitopes derived from tumor-specific antigens that are capable of being recognized by $CD4^+$ T cells in the context of HLA class II molecules, is carried out as follows:

Fifteen-mer peptides overlapping by 10 amino acids, derived from a tumor-specific antigen, are generated using standard procedures. Dendritic cells (DC) are derived from PBMC of a normal donor using GM-CSF and IL-4 by standard protocols. $CD4^+$ T cells are generated from the same donor as the DC using MACS beads (Miltenyi Biotec, Auburn, CA) and negative selection. DC are pulsed overnight with pools of the 15-mer peptides, with each peptide at a final concentration of 0.25 μg/ml. Pulsed DC are washed and plated at 1×10^4 cells/well of 96-well V-bottom plates and purified $CD4^+$ T cells are added at 1×10^5/well. Cultures are supplemented with 60 ng/ml IL-6 and 10 ng/ml IL-12 and incubated at 37°C. Cultures are restimulated as above on a weekly basis using DC generated and pulsed as above as antigen presenting cells, supplemented with 5 ng/ml IL-7 and 10 U/ml IL-2. Following 4 *in vitro* stimulation cycles, resulting $CD4^+$ T cell lines (each line corresponding to one well) are tested for specific proliferation and cytokine production in response to the stimulating pools of peptide with an irrelevant pool of peptides used as a control.

EXAMPLE 17

GENERATION OF TUMOR-SPECIFIC CTL LINES USING IN VITRO WHOLE-GENE PRIMING

Using *in vitro* whole-gene priming with tumor antigen-vaccinia infected DC (see, for example, Yee et al, *The Journal of Immunology 157*(9):4079-86 (1996)), human CTL lines are derived that specifically recognize autologous fibroblasts transduced with a specific tumor antigen, as determined by interferon-γ ELISPOT analysis. Specifically, dendritic cells (DC) are differentiated from monocyte cultures derived from PBMC of normal human donors by growing for five days in RPMI medium containing 10% human serum, 50 ng/ml human GM-CSF and 30 ng/ml human IL-4. Following culture, DC are infected overnight with tumor antigen-recombinant vaccinia virus at a multiplicity of infection (M.O.I) of five, and matured overnight by the addition of 3 μg/ml CD40 ligand. Virus is then inactivated by UV irradiation. CD8+ T cells are isolated using a magnetic bead system, and priming cultures are initiated using standard culture techniques. Cultures are restimulated every 7-10 days using autologous primary fibroblasts retrovirally transduced with previously identified tumor antigens. Following four stimulation cycles, CD8+ T cell lines are identified that specifically produce interferon-γ when stimulated with tumor antigen-transduced autologous fibroblasts. Using a panel of HLA-mismatched B-LCL lines transduced with a vector expressing a tumor antigen, and measuring interferon-γ production by the CTL lines in an ELISPOT assay, the HLA restriction of the CTL lines is determined.

EXAMPLE 18

GENERATION AND CHARACTERIZATION OF ANTI-TUMOR ANTIGEN MONOCLONAL ANTIBODIES

Mouse monoclonal antibodies are raised against *E. coli* derived tumor antigen proteins as follows: Mice are immunized with Complete Freund's Adjuvant (CFA) containing 50 μg recombinant tumor protein, followed by a subsequent intraperitoneal boost with Incomplete Freund's Adjuvant (IFA)

containing 10 μg recombinant protein. Three days prior to removal of the spleens, the mice are immunized intravenously with approximately 50 μg of soluble recombinant protein. The spleen of a mouse with a positive titer to the tumor antigen is removed, and a single-cell suspension made and used for fusion to SP2/O myeloma cells to generate B cell hybridomas. The supernatants from the hybrid clones are tested by ELISA for specificity to recombinant tumor protein, and epitope mapped using peptides that spanned the entire tumor protein sequence. The mAbs are also tested by flow cytometry for their ability to detect tumor protein on the surface of cells stably transfected with the cDNA encoding the tumor protein.

EXAMPLE 19

THE EST EXPRESSION PROFILE OF CLONE 336811 DEMONSTRATES IT IS HIGHLY OVEREXPRESSED IN A VARIETY OF TUMOR TYPES COMPARED TO NORMAL TISSUES

Clone 336811 (SEQ ID NO:656) was found to show some degree of sequence homology to Genebank accession number KIAA1755, the cDNA sequence and polypeptide sequences of which are disclosed in SEQ ID NO:2920 and 2921-2931. Electronic Northern (eNorthern) analysis, the process of which is described in detail in Example 6, revealed that clone KIAA1755 demonstrated an excellent EST profile in a variety of cancerous tissues including anaplastic oligodendroglioma, glioblastoma, and meningioma. Expression of KIAA1755 in anaplastic oligodendrogioma was 10 fold higher in brain tumors than in normal brain. This finding indicates that clone KIAA1755 would be useful as a pan-tumor marker, *e.g.*, for the detection of cancers from a variety of tissue sources.

The expression profile of KIAA1755 is consistent with that of a tumor-associated sequence useful as a target for diagnostic and therapeutic applications for cancers such as pancreatic, endometrium, melanoma, prostate, and fetal liver and spleen cancers.

EXAMPLE 20

REAL-TIME PCR ANALYSIS OF CLONE 182036 IDENTIFIES ITS KIDNEY ASSOCIATED CANCER EXPRESSION PROFILE

Example 3 discusses the generation of an electronic subtraction library using placenta and testis as drivers. From this library clone 182036 (SEQ ID NO:577) or antigen OF1096C was selected for further analysis. When SEQ ID NO:577 was searched against publicly available sequence databases, a second cDNA sequence was found, which is disclosed in SEQ ID NO:2932. Previously, the cDNA specific for clone 182036 was shown to be represented four or more times in tumor specific libraries (see Example 6 for details). This clone was also subjected to eNorthern analysis where the ratio of EST sequences present in all tumor libraries were compared to the total number of EST hits. Clone 182036 was determined to have a ratio of greater than 0.4, indicating that is over-expressed in a variety of cancer types.

Real-Time PCR analysis of OF1096C revealed this sequence was highly over-expressed in 100% of the kidney tumors tested compared to normal kidney and an extensive panel of normal tissues that included, pancreas, liver, stomach, small intestine, colon, bladder, lung, heart, brain, spinal cord, pituitary gland, adrenal gland, thyroid gland, salivary gland, esophagus, trachea, bone marrow, resting PBMC, activated T and B cells, spleen, lymph node, skin, skeletal muscle, testis, placenta, and pooled fetal tissues. This expression profile supports the use of OF1096C as a diagnostic and immunotherapeutic target for kidney tumors.

EXAMPLE 21

MICROARRAY ANALYSIS OF SEQUENCES DERIVED FROM BOTH THE TPS1 AND PPS1 SUBTRACTION LIBRARIES IDENTIFIES ADDITIONAL SEQUENCES HAVING CANCER ASSOCIATED EXPRESSION PROFILES

This example describes the micorarray expression analysis of Oncofetal Chip 1. This chip includes 1248 clones from the TPS1 library (which was described in detail in Example 8) and 288 clones from the PPS1 library

(which was described in detail in Example 9). Oncofetal Chip 1 was probed with probe pairs representing a variety of tumor and normal tissues. Analysis consisted of determining the ratio of the mean hybridization signal for a particular cDNA using two sets of probe pairs. The ratio is a reflection of the over- or under-expression of the cDNA within a probe population. Probe groups were designed to determine cDNA sequences with high differential expression in the "tumor tissue" probe group compared to the "normal tissue" probe group. A threshold for the ratio of over-expression between the tumor tissue probe group and the normal tissue probe group was set at 2.0. Twenty-seven clones identified by this analysis were sequenced and compared to sequences in the publicly available Genebank database. These sequences are disclosed in SEQ ID NOs:2933-2959 and described in detail in Table 12. Based on the tumor associated expression profiles of these sequences, the sequences represent attractive targets for both diagnostic and immunotherapeutic applications.

Table 12

Microarray Analysis of Sequences Identified From the TPS1 and PPS1 Subtraction Libraries

SEQ ID NO.	Clone ID	Ratio	Tumor Mean	Normal Tissue Mean	GenBank
2933	94818	3.53	0.25	0.071	GAGE-5 (U19146)
2934	94819	3.73	0.448	0.12	SOX9 (Z46629)
2935	94820	3.75	0.213	0.057	cDNA FLJ25390 fis, (AK058119)
2936	94821	7.68	0.483	0.063	membrane protein-like protein (U21556)
2937	94822	3.58	0.195	0.054	PDZ-binding kinase (AF189722)
2938	94823	8.22	0.719	0.087	stathmin (X53305)
2939	94824	3.05	0.189	0.062	KIAA1570 protein (AB046790)
2940	94825	4.94	0.381	0.077	stathmin (X53305)
2941	94826	5.3	0.271	0.051	plasma serine protease (protein C) inhibitor (J02639)
2942	94827	4.07	0.221	0.054	cDNA FLJ10007 fis, (AK000869)
2943	94828	3.28	0.405	0.124	plasma serine protease (protein C) inhibitor (J02639)
2944	94829	12.93	1.246	0.096	testis-specific protein Y (TSPY)

SEQ ID NO.	Clone ID	Ratio	Tumor Mean	Normal Tissue Mean	GenBank
					(U58096)
2945	94830	5.43	0.473	0.087	heat shock protein (hsp110 family) (APG-1), (NM_014278)
2946	94831	6.25	0.761	0.122	stathmin (X53305)
2947	94832	3.99	0.519	0.13	IFN-inducible gamma2 protein (X59892)
2948	94833	2.09	0.205	0.098	PRTD-NY2 (AY009106)
2949	94834	3.83	0.445	0.116	stathmin (X53305)
2950	94835	3.77	0.259	0.069	GAGE-5 (U19146)
2951	94836	3.37	0.442	0.131	plasma serine protease (protein C) inhibitor (J02639)
2952	94837	3.54	0.244	0.069	collagen type XVII (COL17A1) gene, 3' UTR, long form (AF136185)
2953	94838	5.58	0.645	0.116	pregnancy-specific beta 1-glycoprotein 4 (PSG4) (U18468)
2954	94839	4.81	0.745	0.155	pregnancy-specific beta-glycoprotein e (M20882)
2955	94840	13.11	1.901	0.145	pregnancy-specific glycoprotein beta-1 (SP1) (M30629)
2956	94842	4.46	0.236	0.053	Novel
2957	94844	5.28	0.724	0.137	pregnancy-specific beta 1-glycoprotein 4 (PSG4) (U18468)
2958	94908	3.03	0.337	0.111	chromosome 20q12-13.2 (AL034418)
2959	94909	8.61	1.086	0.126	pregnancy-specific glycoprotein beta-1 (SP1) (M30629)

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS

What is Claimed:

1. An isolated polynucleotide comprising a sequence selected from the group consisting of:

(a) sequences provided in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959.;

(b) complements of the sequences provided in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959;

(c) sequences consisting of at least 20 contiguous residues of a sequence provided in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959;

(d) sequences that hybridize to a sequence provided in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959, under highly stringent conditions;

(e) sequences having at least 75% identity to a sequence of SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959;

(f) sequences having at least 90% identity to a sequence of SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959; and

(g) degenerate variants of a sequence provided in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959.

2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) sequences encoded by a polynucleotide of claim 1;

(b) sequences having at least 70% identity to a sequence encoded by a polynucleotide of claim 1;

(c) sequences having at least 90% identity to a sequence encoded by a polynucleotide of claim 1;

(d) sequences provided in SEQ ID NOs:2786, 2895-2899, 2902-2903, 2912-2919, and 2921-2931;

(e) sequences having at least 70% identity to sequences provided in SEQ ID NOs: 2786, 2895-2899, 2902-2903, 2912-2919, and 2921-2931; and

(f) sequences having at least 90% identity to a sequence provided in SEQ ID NOs:2786, 2895-2899, 2902-2903, 2912-2919, and 2921-2931.

3. An expression vector comprising a polynucleotide of claim 1 operably linked to an expression control sequence.

4. A host cell transformed or transfected with an expression vector according to claim 3.

5. An isolated antibody, or antigen-binding fragment thereof, that specifically binds to a polypeptide of claim 2.

6. A method for detecting the presence of a cancer in a patient, comprising the steps of:

(a) obtaining a biological sample from the patient;

(b) contacting the biological sample with a binding agent that binds to a polypeptide of claim 2;

(c) detecting in the sample an amount of polypeptide that binds to the binding agent; and

(d) comparing the amount of polypeptide to a predetermined cut-off value and therefrom determining the presence of a cancer in the patient.

7. A fusion protein comprising at least one polypeptide according to claim 2.

8. An oligonucleotide that hybridizes to a sequence recited in SEQ ID NOs: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, and 2933-2959 under highly stringent conditions.

9. A method for stimulating and/or expanding T cells specific for a tumor protein, comprising contacting T cells with at least one component selected from the group consisting of:

(a) polypeptides according to claim 2;

(b) polynucleotides according to claim 1; and

(c) antigen-presenting cells that express a polynucleotide according to claim 1,

under conditions and for a time sufficient to permit the stimulation and/or expansion of T cells.

10. An isolated T cell population, comprising T cells prepared according to the method of claim 9.

11. A composition comprising a first component selected from the group consisting of physiologically acceptable carriers and immunostimulants, and a second component selected from the group consisting of:

(a) polypeptides according to claim 2;

(b) polynucleotides according to claim 1;

(c) antibodies according to claim 5;

(d) fusion proteins according to claim 7;

(e) T cell populations according to claim 10; and

(f) antigen presenting cells that express a polypeptide according to claim 2.

12. A method for stimulating an immune response in a patient, comprising administering to the patient a composition of claim 11.

13. A method for the treatment of cancer in a patient, comprising administering to the patient a composition of claim 11.

14. A method for determining the presence of a cancer in a patient, comprising the steps of:

(a) obtaining a biological sample from the patient;

(b) contacting the biological sample with an oligonucleotide according to claim 8;

(c) detecting in the sample an amount of a polynucleotide that hybridizes to the oligonucleotide; and

(d) compare the amount of polynucleotide that hybridizes to the oligonucleotide to a predetermined cut-off value, and therefrom determining the presence of the cancer in the patient.

15. A diagnostic kit comprising at least one oligonucleotide according to claim 8.

16. A diagnostic kit comprising at least one antibody according to claim 5 and a detection reagent, wherein the detection reagent comprises a reporter group.

17. A method for the treatment of cancer in a patient, comprising the steps of:

(a) incubating CD4+ and/or CD8+ T cells isolated from a patient with at least one component selected from the group consisting of: (i) polypeptides according to claim 2; (ii) polynucleotides according to claim 1; and (iii) antigen presenting cells that express a polypeptide of claim 2, such that T cell proliferate;

(b) administering to the patient an effective amount of the proliferated T cells, and thereby inhibiting the development of a cancer in the patient.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 2002/078516 A3

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number:
PCT/US2002/010421

(22) International Filing Date: 28 March 2002 (28.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/280,255	30 March 2001 (30.03.2001)	US
60/315,563	28 August 2001 (28.08.2001)	US
60/347,313	9 January 2002 (09.01.2002)	US

(71) Applicant *(for all designated States except US)*: **CORIXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WANG, Tongtong** [US/US]; 8049 N.E. 28th Street, Medina, WA 98039 (US). **WANG, Siqing (aka "Steven")** [US/US]; 10145 224th Avenue N.E., Redmond, WA 98053 (US). **BANGUR, Chaitanya, S.** [IN/US]; 2102 N. 105th Street, Apt. J101, Seattle, WA 98133 (US). **GAIGER, Alexander** [AT/US]; 1421 42nd Avenue E., Seattle, WA 98112 (US).

(74) Agent: **CHRISTIANSEN, William, T.**; Seed Intellectual Property Law Group PLLC, Suite 6300, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
18 March 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS AND METHODS FOR THE THERAPY AND DIAGNOSIS OF CANCER

(57) Abstract: Compositions and methods for the therapy and diagnosis of cancer, particularly cancers expressing either C-T or C-P antigens, are disclosed. Illustrative compositions comprise one or more tumor polypeptides, immunogenic portions thereof, polynucleotides that encode such polypeptides, antigen presenting cell that expresses such polypeptides, and T cells that are specific for cells expressing such polypeptides. The disclosed compositions are useful, for example, in the diagnosis, prevention and/or treatment of diseases, particularly cancers expressing either C-T or C-P antigens.

WO 2002/078516 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/10421

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68
US CL : 536/23.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NCBI Entrez database nucleotide sequences, National Center for Biotechnology Information, National Library of Medicine, NIH (Bethesda, MD, USA) Accession number AI126136, GI: 3594650, 28 October 1998, nucleotides 1-419 of reference identical to complement of nucleotides 867-1357 of SEQ ID NO: 2785.	1, 3, 4, 8, 11, 15
X	WO 99/047669 A2 (META-GEN GESELLSCHAFT FUR GENOMFORSCHUNG MBH) 23 September 1999 (23.09.1999), page 146, SEQ ID NO: 56, nucleotides 91-850 are identical to nucleotides 1465-2224 of instant SEQ ID NO: 2785	1, 3, 4, 8, 11, 15
X	US 5,484,726 A (BASSET et al.) 16 January 1996 (16.01.1996), SEQ ID NO: 1, especially nucleotides 1-2212 which share 99.8% identity with nucleotides 14-2224 of instant SEQ ID NO: 2785	1, 3, 4, 8, 11, 15

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 December 2002 (12.12.2002)	14 MAR 2003
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Juliet C. Einsmann Telephone No. (703) 308-0196

Janice Ford for

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/10421

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1, 3, 4, 8, 11, 15, all in part with respect to SEQ ID NO 2785

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group I, claim(s) 1, 3, 4, 8, 11 (in part) and 15, drawn to nucleic acids, vectors, and host cells.

Group II, claim(s) 2, 7, and 11 (in part), drawn to polypeptides.

Group III, claim(s) 5 and 11 (in part), drawn to antibodies.

Group IV, claim(s) 6, drawn to methods of detecting cancer, utilizing polypeptides.

Group V, claim(s) 9-10 and 11 (in part), drawn to methods for stimulating T-cells.

Group VI, claim(s) 12-13, drawn to methods for stimulating immune response.

Group VII, claim(s) 14, drawn to methods for diagnosing cancer using polynucleotides.

Group VIII, claim(s) 17, drawn to methods for treating cancer.

The inventions listed as Groups I-VII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The nucleic acids of group I encompass any fragment of these nucleic acids which would hybridize with any of these nucleic acids or contains 20 nucleotides in common with any one of the recited nucleic acids. Basset et al. (US 5484726) teach an isolated nucleic acid that shares 99.8% local similarity with nucleotides 14-2224 of SEQ ID NO: 2785 disclosed herein (see nucleotides 1-2212 of SEQ ID NO: 1 taught by Basset et al.). Therefore there is no special technical feature which links the instant claims. Furthermore, the specifically recited sequences are all related in that they are disclosed as being related to cancer. This is not a special technical feature, because nucleic acids that are indicative of the presence of cancer and/or differentially expressed in cancer tissue were known in the prior art at the time the invention was made (see for example, US 5484726). Thus, the feature which joins the instant inventions is not a "special technical feature" in accordance with the PCT rule 13.2 because it does not "define a contribution which each of the claimed inventions, considered as a whole, makes ***over the prior art***. (emphasis added)."

The groups comprising polynucleotides, polypeptides, and antibodies are additionally drawn to multiple, distinct products lacking the same or corresponding special technical features. The nucleic acids are composed of nucleotides and function in, e.g., methods of nucleic acid hybridization or amplification. The polypeptides differ in both structure and function from either the nucleic acids. The polypeptides are composed of amino acids linked by peptide bonds and arranged in a complex combination of alpha helices, beta pleated sheets, hydrophobic and hydrophilic domains. The polypeptides also differ in function, e.g., fusion proteins with an enzymatic functions. The antibodies are composed of amino acids linked by peptide bonds, antibodies are glycosylated and their tertiary structure is unique, where four subunits (2 light chains and 2 heavy chains) associated via disulfide bonds into a Y-shaped symmetric dimer. The antibodies function in immunoassays. Accordingly, the products differ structurally and functionally from one another. As products of different sets of Groups differ from each other in structure, function, and effect, they do not belong to a recognized class of chemical compound, or have both a "common property or activity" and a common structure as would be required to show that the inventions are "of a similar nature. "

The different methods have different objectives and require different process steps. The methods of detecting cancer of groups IV and VII have the same goal (to detect cancer) but do not share common process steps as one is directed to the detection of polypeptides and one to the detection of polynucleotides. The additional methods of groups V, VI, and VII have separate goals from the diagnostic methods and employ unique process steps. Thus, the claims are not joined by a common special technical feature. In addition to differences in objectives, effects, and method steps, it is again noted that the claims of the present Groups are not directed to the use or detection of molecules having the same or common special technical feature, for the reasons discussed above.

INTERNATIONAL SEARCH REPORT

PCT/US02/10421

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

For the groups which are drawn to nucleic acids (group I), methods for using nucleic acids (groups VI (in part) and VII), the species are SEQ ID NO: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, 2933-2959.

For the claims which are drawn to polypeptides (group II) or methods for using polypeptides (IV, V, VI (in part)) the species are the polypeptides encoded by SEQ ID NO: 2785, 1-2784, 2787-2894, 2900-2901, 2904-2911, 2920, 2932, 2933-2959. If applicant elects claims to antibodies (III) applicant should elect an appropriate corresponding polypeptide from the listed species.

If applicant elects group VI, applicant should further identify one of the compositions recited in claim 11 for search.

If applicant elects group VIII, applicant should further identify one of options (i), (ii), or (iii) as recited in the claim (identifying the particular sequence desired for search).

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons:

The sequences are patentably distinct because they are unrelated sequences, with different structural, functional, and physiochemical characteristics.

Group I, SEQ ID NO: 2785 is the first named invention and will be searched in accordance with PCT Article 17(3)(a).

Continuation of B. FIELDS SEARCHED Item 3:
searched SEQ ID NO 2785 in US PATS, published US PAT applications, EMBL/GENBANK, GENSEQ, EST databases